DEVELOPMENT AND APPLICATION OF A REDUCED ORDER MATHEMATICAL FRAMEWORK TO UNRAVEL THE COMPLEXITY OF TRAUMA INDUCED COAGULOPATHY

A Dissertation

Presented to the Faculty of the Graduate School

of Cornell University

in Partial Fulfillment of the Requirements for the Degree of

Doctor of Philosophy

by

Adithya Sagar Gurram

May 2017

DEVELOPMENT AND APPLICATION OF A REDUCED ORDER MATHEMATICAL FRAMEWORK TO UNRAVEL THE COMPLEXITY OF TRAUMA INDUCED COAGULOPATHY

Adithya Sagar Gurram, Ph.D.

Cornell University 2017

Trauma is the leading cause of death and disability in United States for both children and adults. In response to trauma, the body unleashes a set of coupled programs that affect the functioning of vascular, immune and autonomous nervous systems. In pathological cases, the integrated output of these programs can result in coagulopathy, systemic inflammatory response syndrome (SIRS), multiple organ dysfunction syndrome (MODS) and potentially even death. Nearly 35%-40% of trauma deaths occur due to uncontrolled hemorrhage resulting from trauma-induced coagulopathy (TIC). TIC also plays an important role in modulating inflammation, organ dysfunction and increased susceptibility to sepsis. Clinical trials for treatment strategies targeting TIC have met with limited success. The interlinked nature of coagulant and inflammatory responses, along with patient specific physiological variability, make the treatment of TIC challenging. Understanding TIC requires an integrated multi-scale modeling framework which describes the relevant biochemical networks within the context of the whole-body. Given the complexity and size, embedding large, non-linear models of biochemical networks into a whole body model creates a significant computational challenge.

Thus an objective of this work is to develop a framework that reduces the complexity of high-dimensional mathematical models. We apply this framework to model biochemical networks that are important in TIC. We first investigate the dynamics of coagulation

and understand the impact of protein C pathway on thrombin generation. Thereafter we use this reduced order modeling technique to model complement and fibrinolysis. We identify targets of therapeutic importance in complement and mechanisms that control clot degradation in fibrinolysis. We show that we can capture the dynamics of these complex but varied systems using the reduced order modeling framework.

In addition, we address the problem of training high-dimensional, non-linear models of biological systems. Traditional gradient based methods often fail due to convergence to a local optima or due to the lack of gradient knowledge. We present a novel optimization method that is based on evolutionary algorithms to obtain near optimal parameters within a limited number of function evaluations. We demonstrate that this method obtains optimal solutions on a wide array of non-linear models, faster than existing meta heuristic methods. Taken together this work provides a methodology to rapidly investigate complex biochemical systems by simplifying the model design and experimentation processes.

BIOGRAPHICAL SKETCH

Adithya Sagar spent his early years running around the lawns of Roorkee and in the historic city of Hyderabad in India. His fascination with biology initially aspired him to consider a career as a neurosurgeon. However this fascination along with a proclivity towards maths and computer science landed him at the Indian Institute of Technology (IIT). For the next four years he dabbled in various ventures, met some wonderful people and tried to understand protein structures with math and machine. After a brief stint as a consultant working with Intellectual Property rights he came back to academia. He started his graduate studies at the Weill Hall in Cornell University trying to decode ultrasound signals that enhance clot degradation. Later he spent about four years in Olin Hall understanding the ”why and how” of blood clots within human body, albeit this time with computational models. He hopes his study will further the understanding of trauma induced coagulopathy and reduce the dependency on gradient based optimization methods.

To my family for all those countless sacrifices. Hope they turn out to be worthwhile.

ACKNOWLEDGEMENTS

The significance of many events is truly understood perhaps in retrospect. Looking back, my journey through graduate school in Ithaca, though trying at times, has been my most profound and enriching experience. There were several individuals who contributed to this experience in various ways whom I hope I can acknowledge.

I would like to start by expressing my immense gratitude to all the members on my thesis committee. First and foremost, I want to convey my deepest and heartfelt gratitude to my advisor Prof. Jeffrey Varner. His deep knowledge, passion towards research, "design" aesthetic and a zen like focus towards building models have influenced me greatly. I am also very thankful for his incredible patience and for nudging me at the right moments to get to my goals. Above all I am quite thankful that he provided me the space and time to grow intellectually. It was an absolute privilege knowing you.

Words would fall short in expressing my thanks and appreciation to Prof. Paulette Clancy. It was always a great pleasure talking to her and receiving her words of immense wisdom and insight. More than being a great mentor she was close to being a family member. Thank you so much.

I am extremely thankful to Prof. Peter Frazier, one of the most creatively brilliant persons that I have known. I am particularly grateful for our stimulating conversations and for exposing me to Stochastic Processes and Bayesian Statistics. I also would like to convey my thanks for taking the trouble to attend my exams remotely despite a hectic schedule.

In addition, I am thankful to a number of faculty at Cornell who influenced me greatly. I am very grateful to Prof. Christine Shoemaker for introducing me to evolutionary algorithms and non-linear optimization. I am deeply appreciative of Prof. Lynden Archer for being a great mentor outside of work. Thanks a lot for your immense support and

guidance in the times of need. I am also very thankful to Prof. Susan Daniel for being a really wonderful director of graduate students; Prof. Abe Stroock and Prof. William Olbricht for convincing me to join the PhD program.

There have been several collaborators on the coagulopathy project. In particular I am quite thankful to our collaborators from University of California, Santa Barbara that include Prof. Linda Petzold, Andri Bezzola, Tie Bo Wu and all the members of Petzold lab. I am also very grateful to Prof. Kathleen Brummel-Ziedins and Dr. Thomas Orfeo from University of Vermont for their terrific insight into *in vitro* experimental protocols concerning coagulation and fibrinolysis. I am immensely thankful to both of them for providing us with clinical data related to ROTEM measurements. I am also quite grateful to the Department of Defense for providing the grants to conduct my research.

I have had the most wonderful set of lab mates at Olin. From the Varner lab, I am tremendously thankful to Joe Wayman, Holly Jensen, Kathy Rogers, Ankit Chandra, Lina Aboulmouna, David Bassen, Mason Minot, Rohaine Hsu, David Dai, Mike Vilkhovoy, Rachel LeCover and Nick Horvath. I quite enjoyed my experience spending entire nights in Olin working with David and Mason on the complement project. Thanks guys. I am quite thankful to Rohaine for her superb desserts and food in times of need. I am very thankful to Mike for being patient enough to let me in whenever I got locked out. Thanks to Ankit for a great first two years in 165. It was a lot of fun. I also want to convey my thanks to Ryan Hoang, Mark Unter and Josef Byrne for being terrific researchers despite their heavy coursework. I wish them the very best of luck.

Life in Ithaca would not have been possible without my friends who were a family and a bedrock of support away from home. For this I am highly indebted to several people: Anil, Rohit, Raghu, Anand for the awesome game nights, cricket at Pleasant Street and trips to the All You Care Eats; Ranjith and Manjari, for all the insane amount of cooking and fun; Prashanth, for being an awesome housemate and for always lending a

helping hand; Anubhab, Neeraj, Binit, Ravi for all the superb hiking trips; Snehashis and Himanshu, for being great companions over the last couple of years; Rahul and Neetu for being a great support. Thanks a lot to Zhenia and Jade for all the fun outside work and to Lingfeng for being a wonderful friend.

Thanks to Anubhab and Senjuti for almost being my parents. Hope Akanksha and I were a good prequel for things to come. Finally, thanks to Akanksha for being there. The journey was simply not possible without you. Hope the next phase will be as exciting as this one was. To my family, I do not have the power to articulate anything that comes close to saying thanks. I will always fall short trying to say it.

To Ammamma, Thathu, Chinnamama, Attha, Mohan, Arun, thanks a lot for your immense belief in me. To Amma, Nanna, Thammu, this is as much yours as it is mine. Thanks for always being there.

TABLE OF CONTENTS

LIST OF TABLES

LIST OF FIGURES

CHAPTER 1

INTRODUCTION

Blood is perhaps the most recognized and studied biological component. Yet the complexity encoded in it through millions of years of evolution still poses fascinating questions. Critical to our understanding of coagulopathy or related disorders are the mechanisms that contribute to hemostasis. There have been a plethora of studies over the last two and half centuries investigating hemostasis. Over this period the fundamental question in hemostasis perhaps has transitioned from how does blood clot to why does blood clot, when it does. Understanding this question is challenging partly due to the definition of a normal hemostatic response. Under what circumstances is hemostasis considered normal? What physiological conditions define it? If a normal hemostatic response is defined, what perturbs this *status quo*? How do we quantify the variability that arises due to these perturbations? To answer these questions we need to unravel the dynamics of a complex biochemical phenomena that is tightly coupled with physiological conditions. This motivates us to come up with a holistic yet detailed framework that can address these questions.

Mathematical modeling is at the core in analyzing the emergent behavior of biological systems. A wide array of mathematical paradigms have been developed that capture varying degree of biological complexity (or detail) [6, 7]. The choice of a modeling paradigm thereby, is dictated by the pertinent biological question, the quality and quantity of available data. On one end of the spectrum lie ordinary differential equation (ODE) based models. These models require detailed mechanistic information and can capture complex dynamics. At other end are a number of qualitative approaches that do not rely on detailed mechanistic information but more on network topology. Examples of such

approaches include graph models, Petri nets, constraint-based modeling and logic models like ones based on Boolean logic [7]. Such qualitative approaches can handle large network structures more easily than ODE based models. However such models are generally inadequate in modeling detailed system dynamics.

Given such constraints, modeling hemostatic response during trauma induced coagulopathy (TIC) poses an interesting challenge. The process of coagulation which is central to hemostasis is an intricate physiochemical phenomena [8, 9]. The kinetics of coagulation have been well studied under *in vitro* conditions. However in a coagulopathic state, abnormal physiological conditions and additional biochemical interactions alter the dynamics of coagulation [10, 11]. These interconnected physiological and biochemical phenomena operate on various scales of time and size. To capture the dynamics of these interconnected phenomena, we require an integrated modeling framework that is detailed and can also account for incomplete or missing data. However to describe such integrated frameworks we rely on a large number of parameters and details of interacting species. Due to incomplete information and computational constraints this is always not feasible. Thus we need a paradigm that can *reduce* our dependency on parametric detail within integrated modeling frameworks.

As a step towards this end, in this thesis, a reduced order modeling approach is described that is able to model the dynamics of complex biochemical networks in sufficient detail with fewer parameters. Traditional ODE models rely on saturated enzyme kinetics like Michaelis-Menten kinetics in describing the network. In the reductionist approach, using transfer function based logical rules helps encompass complex logic with few equations. The choice of species and their number is dictated by the pertinent question. For example, a number of studies have shown the importance of protein C pathway in TIC. Thus when we describe our reduced order model of coagulation we investigated the im-

pact of protein C pathway on the overall coagulation dynamics. In a similar vein alternate reduced order models of coagulation may be constructed based on the relevant question.

1.1 An outline of the thesis

The first paper in this thesis (chapter 2) describes an approach to build a reduced order model of coagulation. The production of thrombin, a key protein in coagulation and fibrinolysis is tightly regulated through a number of positive and negative feed back loops. To model thrombin generation, we survey the literature for existing mathematical frameworks for biochemical networks. Through this survey we realize the need for a new framework that reduces number of parameters yet one that captures overall dynamics of the network. We describe a novel reduced modeling approach and capture thrombin dynamics with five equations and 23 parameters. This is an order of magnitude lesser than previous approaches that required hundreds of parameters. In this study we focus on the regulatory effects of protein C, a key component in coagulopathy and train our model on *in vitro* experimental data with and without protein C. To test the predictive power of the model we use hitherto unseen experimental data for thrombin generation. Remarkably the model predicts thrombin generation with different initial conditions. The model also provides insights into regulatory aspects of protein C and identifies mechanisms that are critical to the formation of a stable clot.

In the next chapter (chapter 3) we use the same framework for metabolic modeling. Through the use of logical rules and traditional kinetic modeling we model complex allosteric patterns. In chapter 4 we model complement, a key network in immune response. C3a, C3b, C5a and C5b are important complement proteins that control both innate and adaptive immune response. These proteins also interact with coagulation, fibrinolysis

and other biological networks [12–15]. C5a, in particular, is a critical protein regulating sepsis that results from TIC. Complement initiation occurs through three pathways that involves formation of complex protein structures [16]. We use logical rules to reduce parameters involved in complement initiation and thereby capture the the dynamics of C3a and C5a formation initiated through the lectin and alternate pathways. In chapter 5 we model fibrinolysis using the same framework. Fibrinolysis follows coagulation wherein the clot formed at the end of coagulation is broken down [17, 18]. The thrombin produced during coagulation catalyzes the formation of a cross linked polymer called fibrin. During fibrinolysis an enzyme called plasmin is activated which cleaves fibrin to fibrin degradation products. Fibrin along with platelets makes up a clot and is the key component in determining the stability and strength of a clot. In a clinical setting, fibrin formation and degradation are measured indirectly by measuring viscoelasticity of blood, thereby determining clot strength. The most common and widely used viscoelastic measure is called Rotational Thromboelastometry (ROTEM) [19, 20]. In this study we used ROTEM data from various patients to train our fibrinolysis models. Thereafter we predicted ROTEM dynamics among patients whose data was not used for training. We surprisingly did well and were also able to isolate key mechanisms involved in fibrinolysis. We finally present a fast optimizer for estimating parameters in large biochemical networks. This technique combines Dynamically Dimensioned Search (DDS) [21] and Particle Swarm Optimization (PSO) [22]. On a model of coagulation with 326 parameters and 5 benchmark problems that include commonly used test function for global optimization and 3 large biochemical networks we outperform all the commonly known metaheuristics. Using this method it is possible to generate near optimal solutions for integrated modeling frameworks with a few expensive objective function evaluations. Taken together these reduced order models provide a building block in the construction of an integrated modeling framework for coagulopathy.

CHAPTER 2

DYNAMIC MODELING OF THE HUMAN COAGULATION CASCADE USING REDUCED ORDER EFFECTIVE KINETIC MODELS

Abstract

In this study [1], we present a novel modeling approach which combines ordinary differential equation (ODE) modeling with logical rules to simulate an archetype biochemical network, the human coagulation cascade. The model consisted of five differential equations augmented with several logical rules describing regulatory connections between model components, and unmodeled interactions in the network. This formulation was more than an order of magnitude smaller than current coagulation models, because many of the mechanistic details of coagulation were encoded as logical rules. We estimated an ensemble of likely model parameters (N = 20) from *in vitro* extrinsic coagulation data sets, with and without inhibitors, by minimizing the residual between model simulations and experimental measurements using particle swarm optimization (PSO). Each parameter set in our ensemble corresponded to a unique particle in the PSO. We then validated the model ensemble using thrombin data sets that were not used during training. The ensemble predicted thrombin trajectories for conditions not used for model training, including thrombin generation for normal and hemophilic coagulation in the presence of platelets (a significant unmodeled component). We then used flux analysis to understand how the network operated in a variety of conditions, and global sensitivity analysis to identify which parameters controlled the performance of the network. Taken together, the hybrid approach produced a surprisingly predictive model given its small size, suggest-

[1] Adapted with permission from Sagar A and Varner JD, "Dynamic Modeling of the Human Coagulation Cascade Using Reduced Order Effective Kinetic Models" (2015) Processes, **3**:178–203

ing the proposed framework could also be used to dynamically model other biochemical networks, including intracellular metabolic networks, gene expression programs or potentially even cell free metabolic systems.

2.1 Introduction

Developing mathematical models of biochemical networks is a significant facet of systems biology. Modeling approaches differ in their degree of detail, where the choice of approach is often determined by prior knowledge, or model requirements [6]. Ordinary differential equation (ODE) models are common tools for modeling biochemical systems because of their ability to capture dynamics and encode mechanism. However, ODE models typically come with difficult (or sometimes impossible) parameter identification problems. For example, Gadkar *et al.*, showed that even with near-perfect information, it was often impossible to identify all the parameters in typical signal transduction models [23]. However, it is not clear whether we actually need precise estimates for all model parameters. Bailey suggested more than a decade ago, that achieving qualitative or even quantitative understanding of biological systems should not require complete structural and parametric knowledge [24]. Since Bailey's complex biology with no parameters hypothesis, Sethna showed that model performance is typically sensitive to only a few parameters, a characteristic seemingly universal to multi-parameter models referred to as *sloppiness* [25]. Thus, reasonable predictions may be possible, despite parametric uncertainty, if a few critical parameters are well-defined. For example, Tasseff et al., showed in a model of Retinoic acid (RA) induced differentiation of HL-60 cells, that correct predictions were possible even when 75% of the parameters were known only to an order of magnitude [26]. Perhaps more importantly, ODE models require significant mecha-

nistic information, thereby limiting their utility in poorly understood systems, or conversely explode in size when considering multiple pathways or subsystems. Toward this challenge, logical modeling is an emerging paradigm that encodes causal relationships between model components using quasi-mechanistic non-linear transfer functions [27]. Logical models are highly flexible, and despite their simplicity, they have captured rich behaviors in a variety of systems important to human health [28–30]. However, modeling complex dynamics with logical models is challenging. Thus, there is an unmet need for a third approach which combines ODEs and logical models, where ODEs could encode mechanistic information, while missing or incomplete mechanistic knowledge can be approximated using a logical approach.

In this study, we developed a hybrid approach which combined ODE modeling with logical rules to model a well studied biochemical network, the human coagulation system. Coagulation is an archetype proteolytic cascade involving both positive and negative feedback [31–33]. Coagulation is mediated by a family proteases in the circulation, called factors and a key group of blood cells, called platelets. The central process in coagulation is the conversion of prothrombin (fII), an inactive coagulation factor, to the master protease thrombin (FIIa). Thrombin generation involves three phases, initiation, amplification and termination [34, 35]. Initiation requires a trigger event, for example vessel injury, which leads to the activation of factor VII (FVIIa). Two converging pathways, the extrinsic and intrinsic cascades, then process and amplify this initial coagulation signal. The extrinsic cascade is generally believed to be the main mechanism of thrombinogenesis in the blood [36–38]. Initially, thrombin is produced upon cleavage of prothrombin by fluid phase activated factor X (FXa), which itself has been activated by Tissue Factor/factor VII (TF/FVIIa) [31]. Picomolar amounts of thrombin then activate the cofactors factors V and VIII (fV and fVIII) and platelets, leading to the formation of the tenase and prothrom-

binase complexes on activated platelets. These complexes amplify the early coagulation signal by further activating FXa, and directly converting prothrombin to thrombin. There are several control points in the cascade that inhibit thrombin formation, and eventually terminate thrombin generation. Tissue Factor Pathway Inhibitor (TFPI) inhibits FXa formation catalyzed by TF/FVIIa, while antithrombin III (ATIII) neutralizes several of the proteases generated during coagulation, including thrombin. Thrombin itself also inadvertently plays a role in its own inhibition; thrombin, through interaction with thrombomodulin, protein C and endothelial cell protein C receptor (EPCR), converts protein C to activated protein C (APC) which attenuates the coagulation response by proteolytic cleavage of fV/FVa and fVIII/FVIIIa. Termination occurs after either prothrombin is consumed, or thrombin formation is neutralized by inhibitors such as APC or ATIII.

Previous coagulation models have typically been formulated as systems of nonlinear ordinary differential equations, using mass action or more complex kinetics, to describe the rates of biochemical conversions [39–43]. Mechanistic ODE coagulation models from our laboratory [44, 45] were built upon the earlier studies of Jones and Mann [46], Hockin *et al.* [47], and later Butenas et al., [48] who developed and then subsequently refined highly mechanistic coagulation models. Recently Papadopoulos and co-workers used a phenomenological mathematical model for thrombin generation [49]. Using a set of four ordinary differential equations they were able to derive an equation for temporal evolution of thrombin generation. However the model has a greater focus on thrombin generation and lesser emphasis on the influence of various regulatory elements on the dynamics of coagulation. Unlike our hybrid approach it does not model regulatory detail and thus may be unable to capture the roles of primary inhibitors like ATIII or protein C. Thus in cases where we need to understand the dynamics of coagulation without complex modeling or when there is incomplete mechanistic knowledge, the hybrid approach is a

better candidate. The model also requires that the parameters are adjusted or fine tuned for each data set unlike the hybrid approach where we use a single ensemble across all experiments. However platelet modeling is a significant advantage of this model over the hybrid approach. Future versions of the hybrid model can involve modeling platelets as done by Papdopoulos *et al.*, in their phenomenological model. Other laboratories have also expanded upon Hockin *et al.*, for example by exploring the intrinsic pathway, the role of stochastic fluctuations in coagulation [50], and the dynamics of thrombin mediated clot formation [51]. Other aspects of coagulation have also been modeled, such as platelet biochemistry [52], multi-scale models of clot formation [53, 54], and transport inside clots [55]. However, these previous studies were largely based upon extensive mechanistic knowledge. This is possible because blood, while enormously complex, can be systematically interrogated. Other systems, such as intracellular signaling networks, are much more difficult to experimentally interrogate. Towards this unmet need, we formulated a hybrid modeling approach which combines ODEs and logical rules to model biochemical processes for which a complete mechanistic understanding is missing. We tested this approach by modeling the human coagulation cascade. The hybrid model consisted of only five differential equations augmented with several logical rules. Thus, the model was more than an order of magnitude smaller than comparable purely ODE models in the literature. We estimated the model parameters from *in vitro* extrinsic coagulation data sets, in the presence of ATIII, with and without the protein C pathway. We then compared the model predictions with thrombin data sets, for both normal and hemophilic coagulation, that were not used for model training. Once validated, we performed flux and sensitivity analysis on the model to estimate which parameters were critical to model performance in several conditions. The reduced order hybrid approach produced a surprisingly predictive coagulation model, suggesting this framework could potentially be used to model other biochemical networks important to human health.

2.2 Results

2.2.1 Formulation of Reduced Order Coagulation Models

We developed a reduced order extrinsic coagulation model to test our hybrid modeling approach (Fig. 2.1). The core of our model was based upon the earlier work of Ismagilov and coworkers [56–59], where we added initiation, factor dependence, and specific inhibition terms to the earlier simplified model. A trigger event initiates thrombin formation (FIIa) from prothrombin (fII) through a lumped initiation step. This step loosely represents the initial activation of thrombin by activated FXa. Once activated, thrombin catalyzes its own formation (amplification step), and inhibition via the conversion of protein C to activated protein C (APC). Antithrombin III (ATIII) inhibits amplification, while APC and tissue factor pathway inhibitor (TFPI) potentially inhibit both initiation and amplification. All initiation and inhibition processes, as well as the dependence of amplification upon other coagulation factors, was approximated using our rule-based approach (Fig. 3.2). Individual regulatory contributions to the activity of pathway enzymes were integrated into control coefficients (ν's) using an integration rule (min/max). These control coefficients then modified the rates of model processes at each time step. Hill-like transfer functions $0 \leq f(\mathcal{Z}) \leq 1$ quantified the contribution of components upon a target process. Components were either individual inhibitor or activator levels or some function of levels, e.g., the product of factor levels. In this study, $\mathcal{Z}$ corresponded to the abundance of individual inhibitors or activators, with the exception of the dependence of amplification upon specific coagulation factors (modeled as the product of factors). When a process was potentially sensitive to multiple inputs, logical integration rules were used to select which transfer functions influenced the process at any given time. In our proof

of concept model, we used a winner takes all strategy; the maximum or minimum transfer function was selected at any given time step. However, other integration rules are certainly possible. Taken together, while the reduced order coagulation model encodes significant biological complexity, it is highly compact (consisting of only five differential equations). Thus, it will serve as an excellent proof of principle example to study the reduction of a highly complex human subsystem.

2.2.2 Identification of Model Parameters Using Particle Swarm Optimization

A critical challenge for any dynamic model is the estimation of kinetic parameters. We estimated kinetic and control parameters simultaneously from eight *in vitro* time-series coagulation data sets with and without the protein C pathway. The residual between model simulations and experimental measurements was minimized using particle swarm optimization (PSO). A population of particles (N = 20) was initialized with randomized kinetic and control parameters and allowed to search for parameter vectors that minimized the residual. However, not all parameters were varied simultaneously. We partitioned the parameter estimation problem into two subproblems based upon the biological organization of the training data; (i) estimation of parameters associated with thrombin formation in the absence of the protein C pathway and (ii) estimation of parameters associated with the protein C pathway. Only those parameters associated with each subproblem were varied during the optimization procedure for that subproblem, e.g., thrombin parameters were *not* varied during the protein C subproblem.The PSO procedure was run for 20 generations for each subproblem, where each generation was 1200 iterations. The best particle from each generation was used to generate the particle population for the next

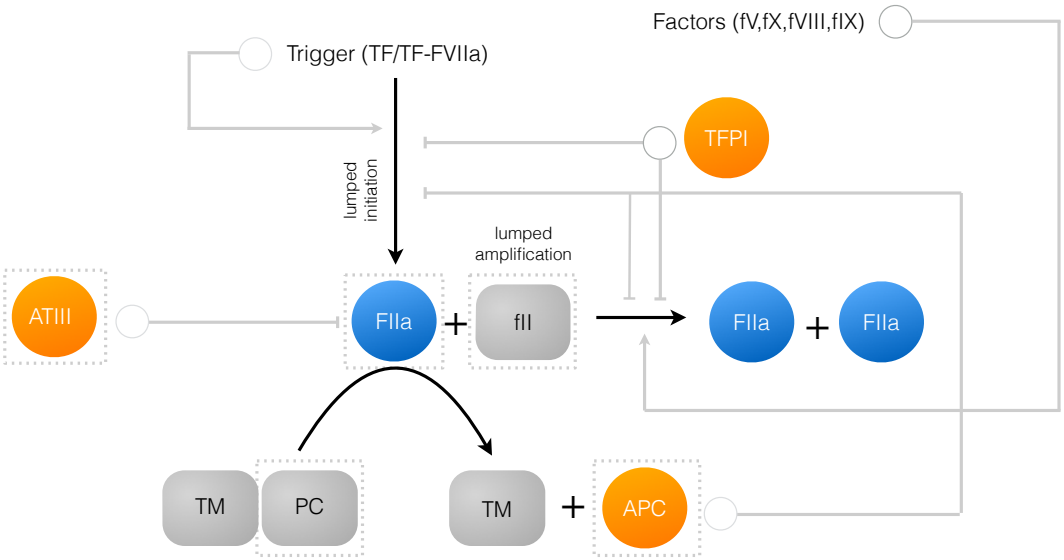

Figure 2.1: Schematic of the connectivity of the reduced order coagulation model. A trigger compound, e.g., TF/FVIIa initiates thrombin production (FIIa) from prothrombin (fII). Once activated, thrombin catalyzes its own activation (amplification step), as well as its own inhibition via the conversion of protein C to activated protein C (APC). APC and tissue factor pathway inhibitor (TFPI) inhibit initiation and amplification, while antithrombin III (ATIII) directly inhibits thrombin. All inhibition steps and trigger-induced initiation were modeled using a rule-based approach. Likewise, the dependence of amplification on other coagulation factors was also modeled using a rule-based approach. The abundance of the highlighted species (in the dashed boxes) was governed by an ordinary differential equation. All other species were assumed to be constant.

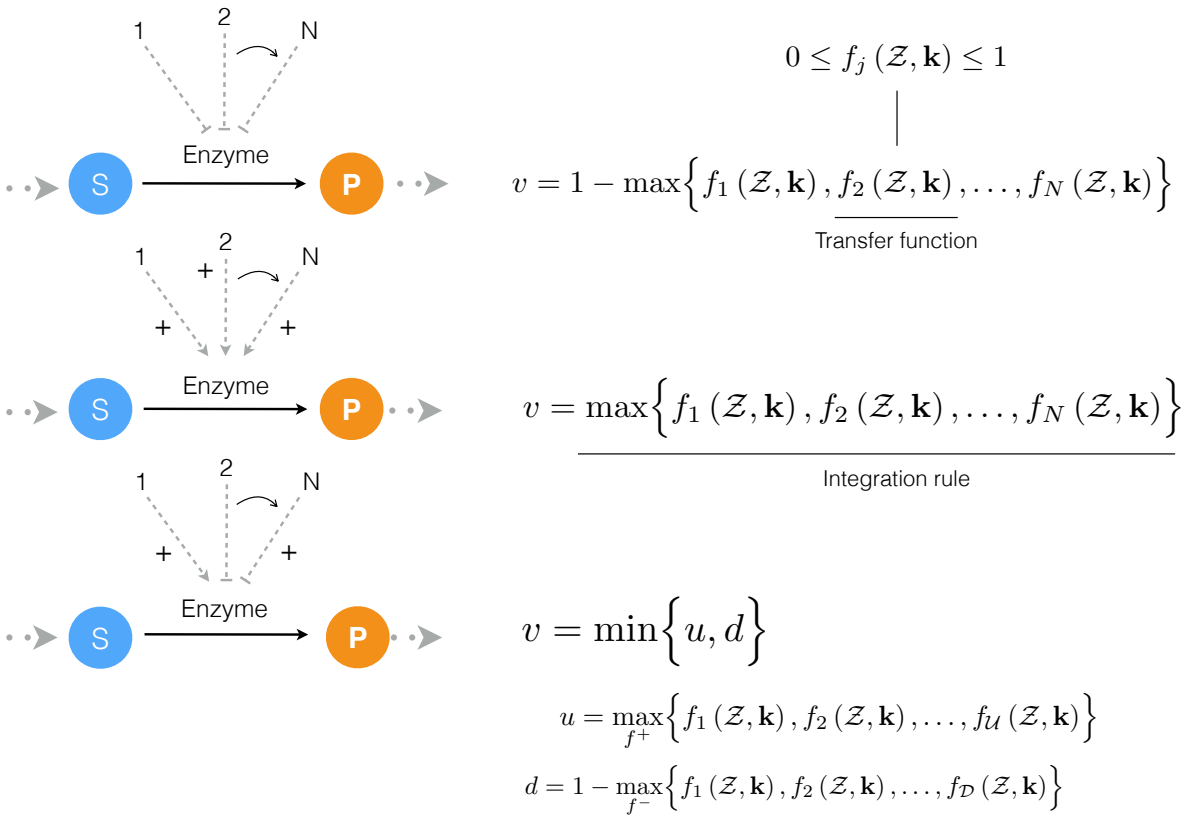

Figure 2.2: Schematic of rule-based effective control laws. Traditional enzyme kinetic expressions, e.g., Michaelis-Menten or multiple saturation kinetics are multiplied by an enzyme activity control variable $0 \leq v_j \leq 1$. Control variables are functions of many possible regulatory factors encoded by arbitrary transfer functions of the form $0 \leq f_j(\mathcal{Z}) \leq 1$. At each simulation time step, the v_j variables are calculated by evaluating integration rules such as the max or min of the set of transfer functions $f_1, \ldots, f_n$ influencing the activity of enzyme E_j.

generation. We rotated the subproblems, starting with subproblem 1 in the first generation.

The experimental training data for parameter estimation was reproduced from the experiments of Butenas and co-workers [1]. In these experiments thrombin generation was initiated by FVIIa-TF using mean plasma concentrations of coagulation proteins and inhibitors. To prepare FVIIa-TF, TF (0.5 nmol/L) was relipidated into 400 μmol/L of phospholipid vesicles (PCPS) by incubation in 20 mmol/L HEPES, 150 mmol/L NaCl, and 2 mmol/L $CaCl_2$ pH 7.4 (HBS/Ca^{2+}) for 30 minutes at 37 °C. The relipidated TF was incubated with 10 pmol/L factor VIIa for 20 minutes to allow the formation of FVIIa-TF. Factors V, VIII and thrombomodulin (Tm) (when protein C activation is required) were added to FVIIa-TF complex. Thrombin generation was then initiated by adding equal volumes of this mixture with a mixture containing prothrombin, factor IX and factor X, TFPI, AT-III and protein C (added when required), protein S (added when required) and factor XI (added when required). In the experimental training data sets that we used for parameter estimation 5 pmol/L FVIIa-TF was used along with 200 μmol/L of phospholipid vesicles (PCPS) to initiate thrombin generation. When protein C pathway was involved, protein C and protein S were at mean plasma concentrations and 0.1 nmol/L Tm was used. All the other coagulation proteins and inhibitors i.e. factors X, IX, V, and VIII, prothrombin, TFPI and AT-III were at their mean plasma concentration levels.

The reduced order coagulation model captured the role of initial prothrombin abundance, and the decay of the thrombin signal following from ATIII activity (Fig. 2.3). However, we systematically under-predicted the thrombin peak and the strength of ATIII inhibition in this training data set. On the other hand, with fixed thrombin parameters, we captured peak thrombin values and the decay of the thrombin signal (at least for the 150% fII case) in the presence of both ATIII and the protein C pathway (Fig. 2.4). Lastly,

we were unable to capture global differences in initiation time *across* separate data sets with a single ensemble of model parameters. These differences likely resulted from normal experimental variability. For example, different thrombin generation experiments within the training data (at the same physiological factor levels) had significantly different initiation times (data not shown). However, the inability to globally capture initiation time also highlighted a potential shortcoming of the initiation module within the model. To capture the variability in initiation time *across* training data sets, we included a constant time-delay parameter (T_D) for each data group. The delay parameter was constant within a data set, but allowed to vary *across* training data sets. Introduction of the delay parameter allowed the model to simulate multiple training data sets using a single ensemble of model parameters. Taken together, the model identification results suggested that our hybrid approach could reproduce a panel of thrombin generation data sets in the neighborhood of physiological factor and inhibitor concentrations. However, it was unclear whether the reduced order model could predict new data, without updating the model parameters.

2.2.3 Validation of the Reduced Order Coagulation Model

We tested the predictive power of the reduced order coagulation model with validation data sets not used during model training. Two validation data sets were used, thrombin generation for various prothrombin and ATIII concentrations with the protein C pathway, and thrombin generation in normal versus hemophilic plasma in the presence of the protein C pathway. Lastly, we compared the qualitative output of the model to rFVIIa addition in the presence of hemophilia. The hemophilia case was an especially difficult test as it was taken from a different study which used a plasma-based *in vitro* assay involving

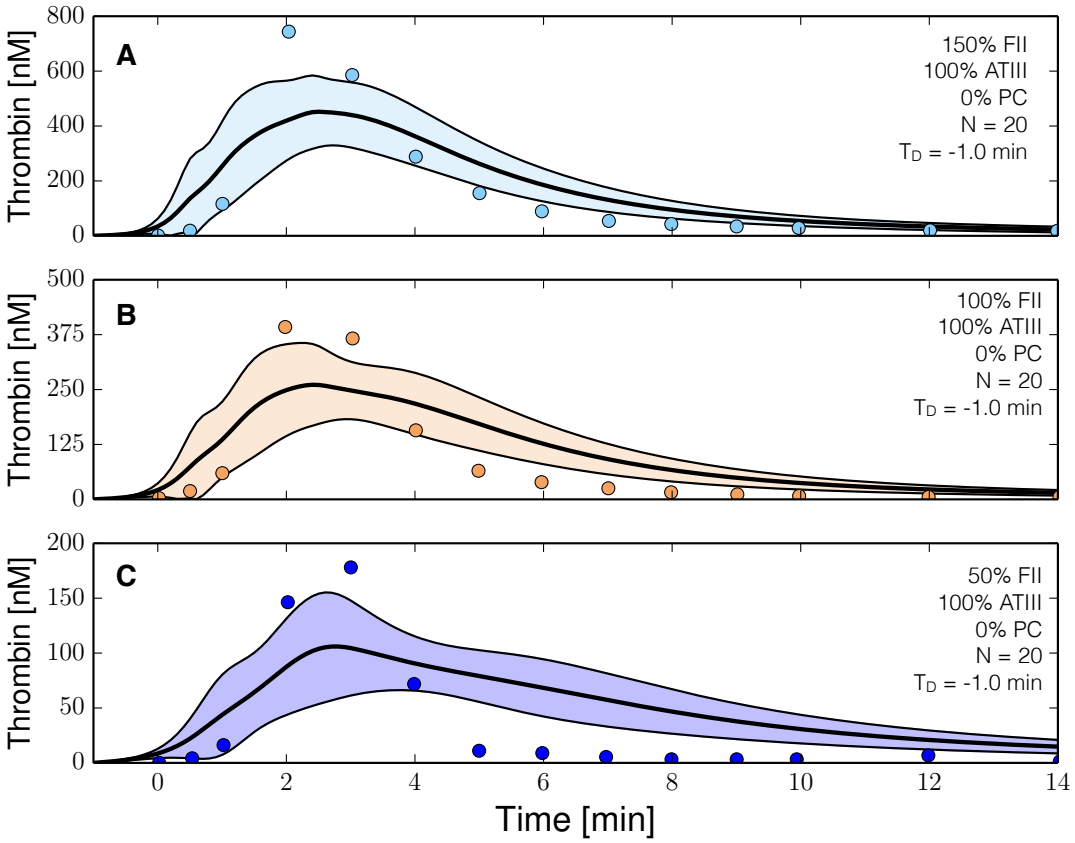

Figure 2.3: Reduced order coagulation model training simulations. Reduced order coagulation model parameters were estimated using particle swarm optimization (PSO) without the protein C pathway as a function of prothrombin. Solid lines denote the simulated mean value of the thrombin profile for N = 20 independent particles, points denote experimental data. The shaded region denotes the 99% confidence estimate of the mean simulated thrombin value (uncertainty in the model simulation). (A,B,C) depict training data and results for 150%, 100% and 50% of physiological prothrombin levels in the absence of protein C pathway. The experimental training data was reproduced from the study of Butenas *et al.* [1][2]. All factors and control proteins in these experiments were at their physiological concentration unless otherwise denoted.

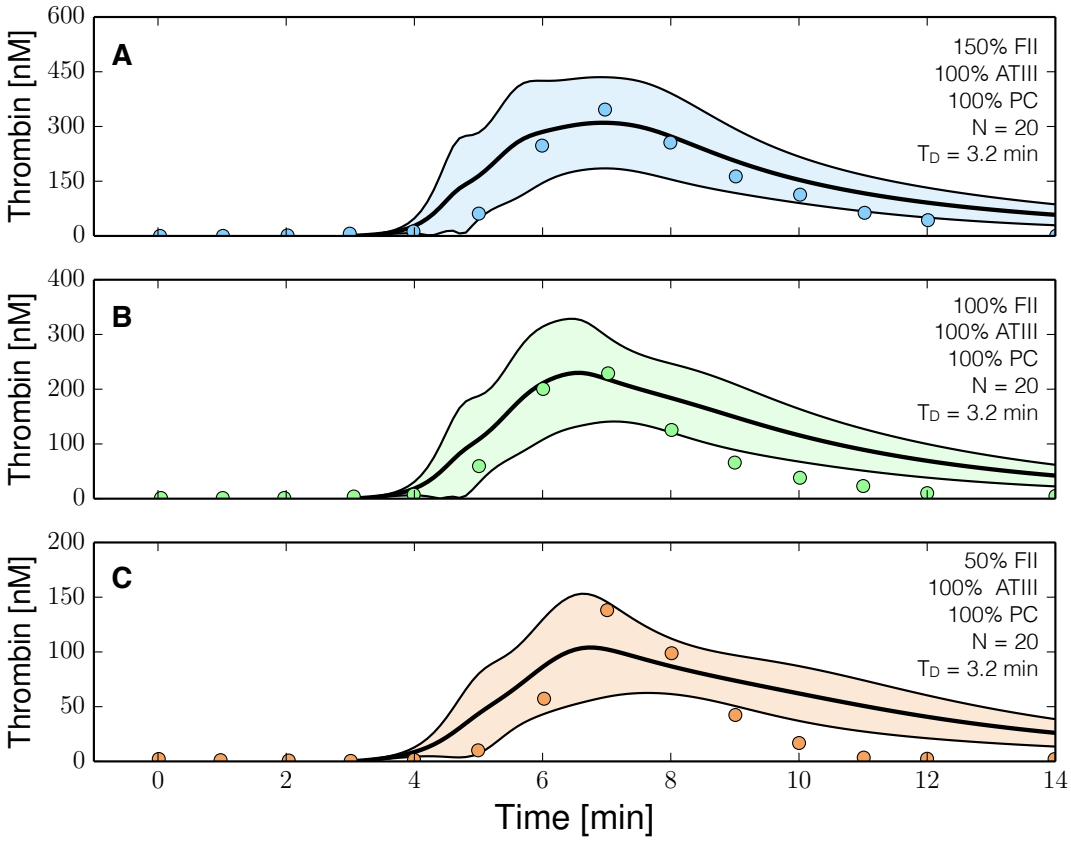

Figure 2.4: Reduced order coagulation model training simulations. Reduced order coagulation model parameters were estimated using particle swarm optimization (PSO) with the protein C pathway as a function of prothrombin. Only APC pathway parameters were allowed to vary in these simulations keeping the parameters estimated without protein C pathways constant. Solid lines denote the simulated mean value of the thrombin profile for N = 20 independent particles, points denote experimental data. The shaded region denotes the 99% confidence estimate of the mean simulated thrombin value (uncertainty in the model simulation). (A,B,C) depict training data and results for 150%, 100% and 50% of physiological prothrombin levels in the presence of the protein C pathway. The experimental training data was reproduced from the study of Butenas *et al.* [1][2]. All factors and control proteins in these experiments were at their physiological concentration unless otherwise denoted.

platelets instead of phospholipid vesicles (PCPS). All kinetic and control parameters were fixed for the validation simulations. The only globally adjustable parameter T_D, was fixed within each validation data set but allowed to vary between data sets. The reduced order model predicted the thrombin generation profile for ratios of prothrombin and ATIII in the absence of the protein C pathway (Fig. 2.5). Simulations near the physiological range (fII,ATIII) = (100%, 100%) or (125%,75%) tracked the measured thrombin values (Fig. 2.5B and C). On the other hand, predictions for factor levels outside of the physiological range (fII,ATIII) = (50%, 150%) or (150%, 50%), while qualitatively consistent with measured thrombin values, did show significant deviation from the measurements (Fig. 2.5A and D). Likewise, simulations of thrombin generation in normal versus hemophilia (missing both fVIII and fIX) were consistent with measured thrombin values (Fig. 2.6). We modeled the dependence of thrombin amplification on factor levels using a product rule ($\mathcal{Z} = fV \times fX \times fVIII \times fIX$), which was then was integrated using a min integration rule into the control variable governing amplification. Thus, in the absence of fVIII or fIX, the amplification control variable evaluated to zero, and the only thrombin produced was from initiation (Fig. 2.6B). However, the decay of the thrombin signal was underpredicted in the normal case (Fig. 2.6A), while the activated thrombin level was overpredicted in hemophilia simulations, although thrombin generation was far less than normal (Fig. 2.6B). Taken together, the reduced order model performed well in the physiological range of factors, even with unmodeled components such as platelet activation in the hemophilia data set.

The model ensemble predicted a direct correlation between thrombin generation and rFVIIa addition in hemophilia (Fig. 2.7). In the current model, we cannot distinguish between different initiation sources, e.g., TF/FVIIa versus rFVIIa, as we have only a single lumped initiation source (trigger). Thus, we simulated the addition of rFVIIa in

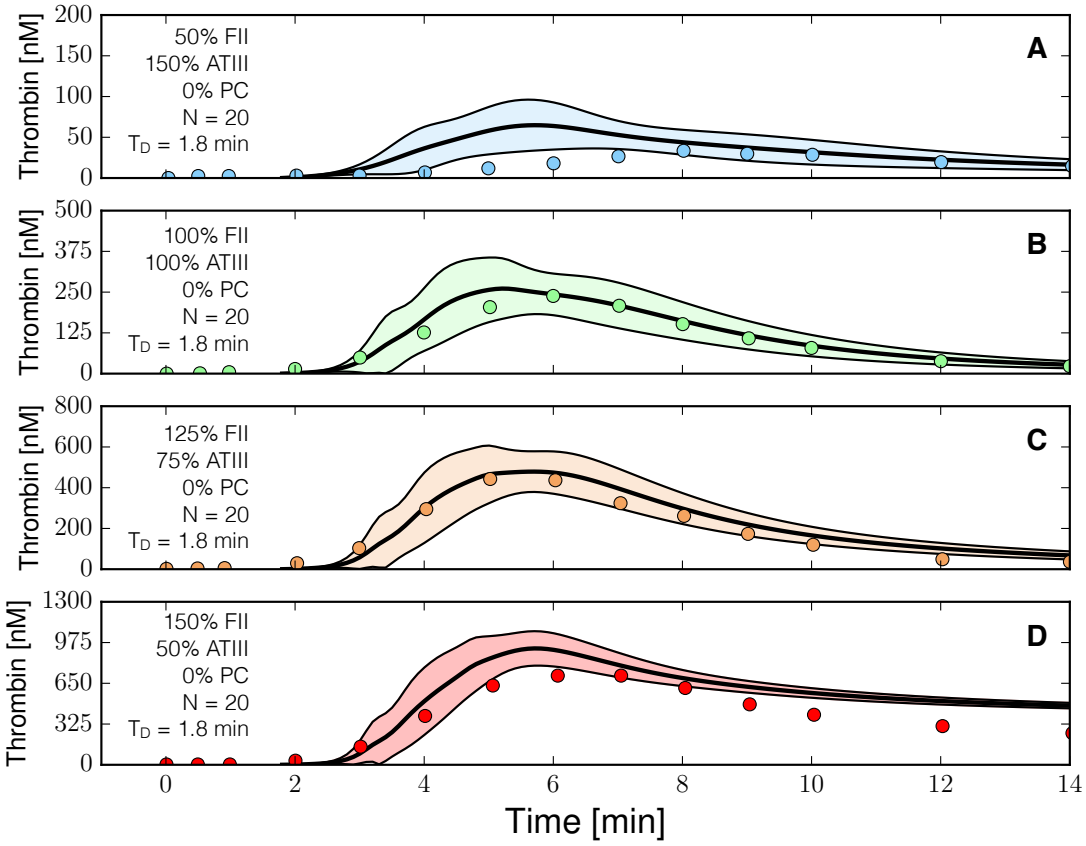

Figure 2.5: Reduced order coagulation model predictions versus experimental data for normal coagulation. The reduced order coagulation model parameter estimates were tested against data not used during model training. Simulations of different levels of prothrombin and ATIII were compared with experimental data in the absence of the protein C pathway. Solid lines denote the simulated mean value of the thrombin profile for N = 20 independent particles, points denote experimental data. The shaded region denotes the 99% confidence estimate of the mean simulated thrombin value (uncertainty in the model simulation). (A,B,C,D) prediction results for (FII,ATIII): (50%,150%), (100%, 100%), (125%, 75%) and (150%, 50%) of physiological prothrombin and ATIII levels in the absence of the protein C pathway. The experimental validation data was reproduced from the study of Butenas *et al.* [1][2]. All factors and control proteins were at their physiological concentration unless otherswise denoted.

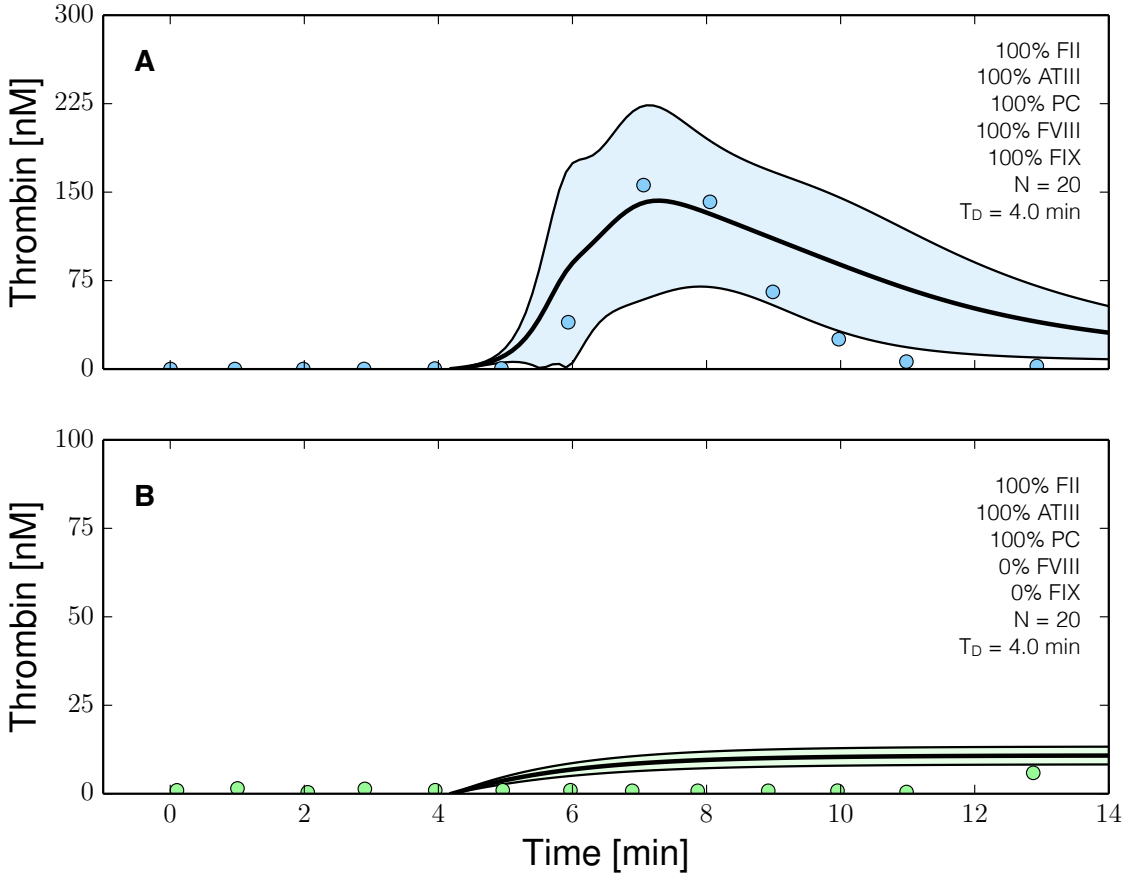

Figure 2.6: Reduced order coagulation model predictions versus experimental data with and without FVIII and FIX. The reduced order coagulation model parameter estimates were tested against data not used during model training. Simulations of normal thrombin formation with ATIII and the protein C pathway were compared with thrombin formation in the absence of fVIII and fIX. Solid lines denote the simulated mean value of the thrombin profile for N = 20 independent particles, points denote experimental data. The shaded region denotes the 99% confidence estimate of the mean simulated thrombin value (uncertainty in the model simulation). (A,B) prediction results for normal thrombin generation and thrombin generation in hemophilia. All factors and control proteins were at their physiological concentration unless others noted. The experimental validation data was reproduced from the study of Allen *et al.* [3].

hemophilia by removing fVIII and fIX from the model, and modulating the initial level of trigger. Simulations with a baseline level of trigger were consistent with the previous hemophilia simulations, where the only thrombin produced was from initiation (Fig. 2.7A, 1× trigger). However, as we increased the trigger strength, the thrombin profile began to approximate normal coagulation, showing a pronounced peak albeit with a slower peak time (Fig. 2.7B, $t^{**} > t^{*}$). Further increases in trigger strength resulted in decreased thrombin peak time and increased maximum thrombin values (Fig. 2.7A, 50× trigger). Thus, for large trigger values (200×trigger), the hemophilic thrombin profile approximated normal coagulation, where peak thrombin was achieved shortly after administration and 95% of the thrombin was gone by 20 min after initiation. We performed flux analysis to understand how the reduced order coagulation model balanced initiation, amplification and termination of thrombin generation for normal and hemophilic coagulation. Analysis of the reaction flux through the reduced order network for thrombin generation in normal, hemophilia and rFVIIa-treated hemophilia identified three distinct operational modes (Fig. 2.8). We calculated the flux through four lumped reactions, initiation, amplification, thrombin-induced APC generation and total thrombin inhibition (including both APC and ATIII action). Directly after the addition of a trigger (e.g., TF/FVIIa or rFVIIa), the lumped initiation flux was the largest for all three cases. However, within a few minutes enough thrombin was generated by the initiation mechanism to induce the amplification stage. During amplification, thrombin catalyzes its own formation and inhibition by generating activated protein C (APC), a potent inhibitor of the coagulation cascade. For normal coagulation, amplification and thrombin inhibition were the dominate reactions by 6 min after initiation (Fig. 2.8, left). After 10 min, the dominate reaction had shifted to thrombin inhibition (both ATIII and APC action). In hemophilia (missing both fVIII and fIX), the amplification reaction did not occur, and thrombin was produced only by initiation (Fig. 2.8, center). Initiation was quickly inhibited by APC,

and the thrombin level stabilized (eventually decaying at longer times because of ATIII activity). Lastly, when 50×trigger was used to induce thrombin formation in hemophilia (absence of fVIII/fIX), initiation mechanisms dominated for up to 6 min following initiation (Fig. 2.8, right). Similar to hemophilia alone, no amplification occurred in the 50×trigger+hemophilia case, and the rate of thrombin generation was extinguished by the combined action of ATIII and APC. Taken together, the hybrid modeling approach captured the transition between the modes of thrombin generation, as well as the role that inhibitors play in attenuating the thrombin generation rate. Thus, the transfer function approach encoded the inhibitory logic of this cascade in the absence of specific mechanism.

2.2.4 Global Sensitivity Analysis of the Reduced Order Coagulation Model

We conducted a global sensitivity analysis to estimate which parameters controlled the performance of the reduced order model. We calculated the sensitivity of the time to maximum thrombin (peak time) and the thrombin exposure (area under the thrombin curve) for different levels of prothrombin, and protein C (Fig. 2.9). Globally, 41% of the parameters shifted in importance between the (fII,PC) = (50%, 0%) and (150%,100%) cases for the peak thrombin time (Fig. 2.9A). The majority of these shifts involved the interaction between increased prothrombin and the protein C pathway, while only 5% were directly associated with increased prothrombin alone. The rate constant for thrombin amplification was the most important parameter controlling the peak thrombin time. While this parameter was differentially important for different prothrombin levels, and in the presence or absence of the activated protein C pathway, it was consistently the most

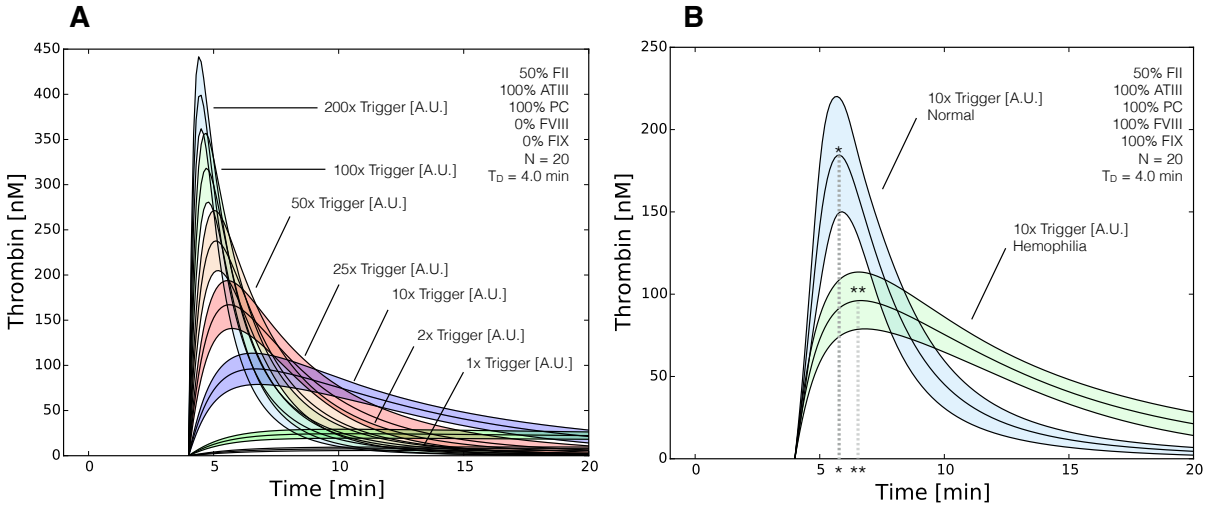

Figure 2.7: Reduced order coagulation model predictions of rFVIIa administration. A: Simulations of thrombin formation in the presence of ATIII and the protein C pathway were conducted for a range of trigger values (1x - 200x nominal) in the absence of fVIII and fIX. B: Comparison of thrombin generation for normal versus hemophilia for 10x nominal trigger. Solid lines denote the simulated mean value of the thrombin profile for N = 20 independent particles. The peak thrombin time for normal coagulation (t^*) is less than rFVIIa induced coagulation in hemophilia (t^{**}), while the peak thrombin value was greater in normal coagulation. The shaded region denotes the 99% confidence estimate of the mean thrombin value (uncertainty in the model simulation). All factors and control proteins were at their physiological concentration unless others noted.

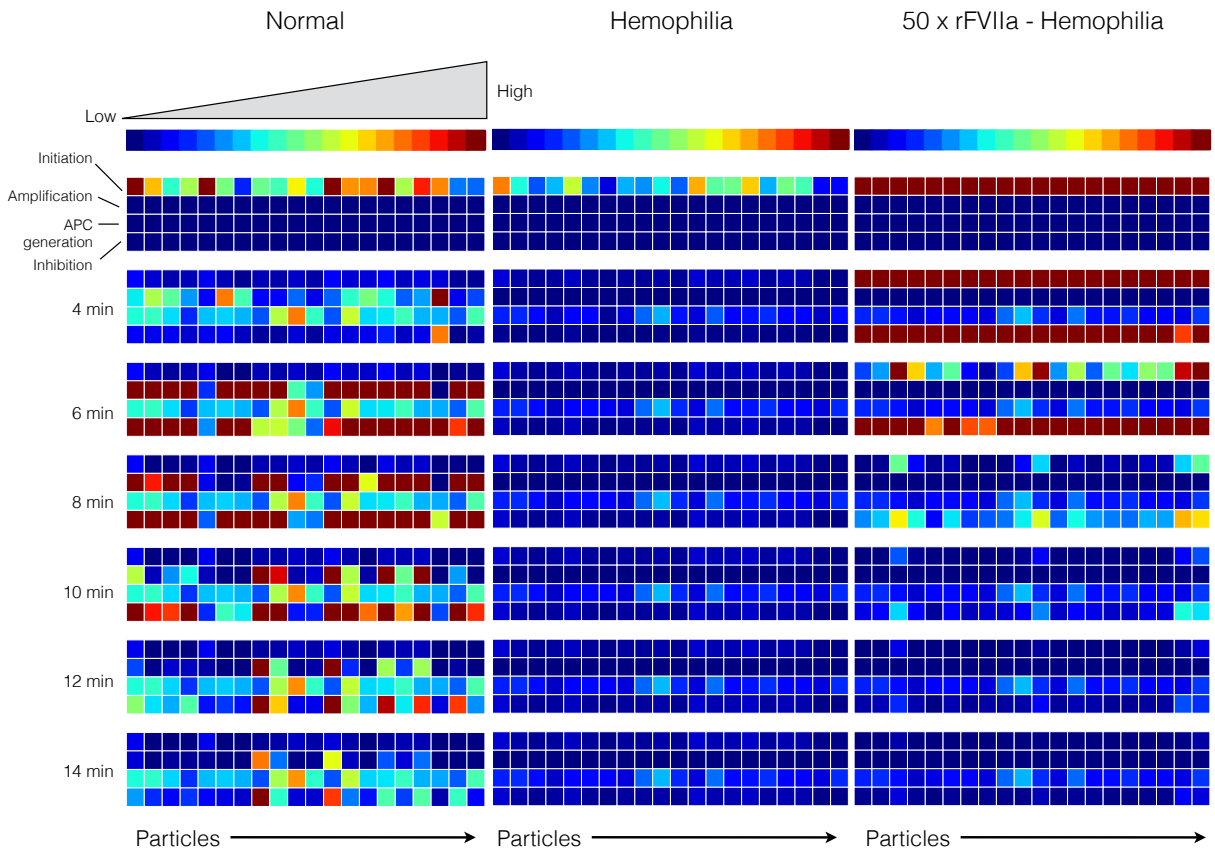

Figure 2.8: Reaction flux distribution as a function of time for thrombin generation under normal (left), hemophilia (center) and rFVIIa treated hemophilia (right). Reaction flux was calculated for each particle at $T = 0, 4, 6, 8, 10, 12, 14$ min after the initiation of coagulation. Reaction fluxes were calculated for each particle in the parameter ensemble (N = 20). Blue colors denote low flux values while red colors denote high flux values.

sensitive parameter in the model. The saturation constant governing thrombin amplification was the second most important parameter, followed by the initiation control gain parameter. Other important parameters influencing the thrombin peak time included the control gain for activated protein C formation, and the rate constant controlling ATIII inhibition of thrombin activity. On the other hand, only 27% of the model parameters were differentially sensitive between the (fII,PC) = (50%, 0%) and (150%,100%) cases for thrombin exposure (Fig. 2.9B). Of these parameters, all of the shifts were associated with the interplay between thrombin formation and the protein C pathway. The rate constant controlling ATIII inhibition was the most important parameter controlling the thrombin exposure. While this parameter was less important in the presence of protein C for 150% prothrombin levels, it was significantly above all other parameters. Similar to the peak time, for 150% prothrombin, the control gain for activated protein C formation was differentially important along with the rate constant controlling amplification. However, the amplification parameter was much less important for thrombin exposure versus peak time.

2.3 Discussion

In this study, we developed a reduced order model of the human coagulation cascade. We modeled coagulation because it is well studied, has a complex architecture, and has an abundance of experimental data available for model identification and validation. However, coagulation was just a proof of concept test of our approach. The proposed hybrid framework could also be used to dynamically model other biochemical networks, including intracellular metabolic networks, gene expression programs or potentially even cell free metabolic systems. The model consisted of five differential equations augmented

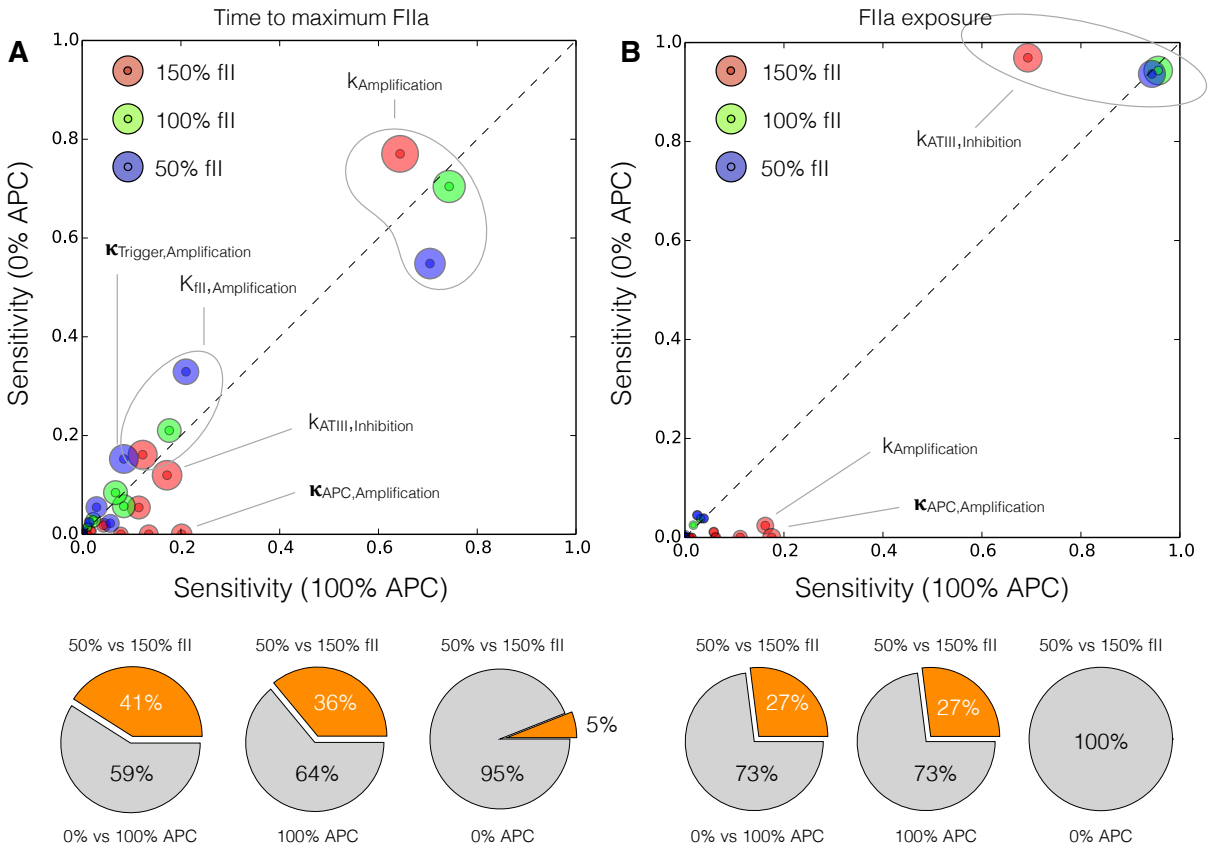

Figure 2.9: Global sensitivity analysis of the reduced order coagulation model with respect to the model parameters. A: Sensitivity analysis of the thrombin peak time for different prothrombin levels (150%,100% and 50% of the physiological value) as a function of activated protein C. B: Sensitivity analysis of the thrombin exposure for different prothrombin levels (150%,100% and 50% of the physiological value) as a function of activated protein C. Points denote the mean total sensitivity value, while the area around each point denotes the uncertainty in the sensitivity value. The gray dashed line denotes the 45° degree diagonal, if sensitivity values are equal for different conditions they will lie on the diagonal. Sensitivity values significantly above or below the diagonal indicate differentially important model parameters. The radius of the shaded region around each total sensitivity value was the maximum uncertainty in that value estimated by the Sobol method.

with several logical rules describing regulatory connections between model components and unmodeled interactions in the network. We estimated model parameters from *in vitro* extrinsic coagulation data sets, in the presence of ATIII, with and without the protein C pathway. To estimate parameters, the residual between model simulations and experimental measurements was minimized using particle swarm optimization (PSO). However, not all of the model parameters were uniquely identifiable, given the training data. Instead, we estimated an ensemble of likely parameter sets (N = 20) from eight *in vitro* time-series coagulation data sets with and without the protein C pathway. Ensemble approaches have been used previously for other signal transduction models [60–64], and for metabolic models [65] to estimate the impact of poorly constrained parameter values or poorly understood network structure on simulation performance. Thus, ensemble approaches are common in the dynamic modeling community. However, a unique feature of the current study is the direct connection between our particle swarm approach, and the parameter ensemble; each particle in our swarm uniquely corresponded to a parameter set in our ensemble. Thus, by constraining particles to operate in different parameter regions, giving each particle a different parameter combination to explore, or perhaps even suppling a different model formulation to each particle we can effectively traverse through complex parameter and model spaces. We validated the ensemble using thrombin data sets taken from multiple laboratories for a variety of experimental conditions not used during training. The ensemble predicted thrombin trajectories for conditions not used for model training, including thrombin generation for normal and hemophilic coagulation in the presence of platelets (a significant unmodeled component). We then used flux analysis to understand how the network operated in a variety of conditions, and global sensitivity analysis to identify which parameters controlled the performance of the network. Flux analysis showed the logical rules formulation encoded the transitions between initiation, amplification and termination of thrombin generation. Sensi-

tivity analysis suggested that the amplification rate constant was more important to the time to peak thrombin, while the ATIII inhibition constant controlled thrombin exposure. Taken together, the proposed hybrid framework produced a surprisingly predictive model, suggesting this approach could be used to effectively model other biochemical networks important to human health.

Malfunctions in coagulation can have potentially fatal consequences. Aggressive clotting involved with Coronary Artery Diseases (CADs), collectively accounts for 38% of all deaths in North America [66]. Coagulation management during surgery can also be challenging, particularly with the increase in clinical use of antithrombotic drugs [67]. Insufficient coagulation due to genetic disorders such as hemophilia can also result in recurrent bleeding. The coagulation factors VIII (fVIII) and IX (fIX) are deficient in Hemophilia A and B, respectively [68–70]. People with mild hemophilia have 5-40% of the normal clotting factor levels while severe hemophiliacs have <1% [70]. Hemophilia can be controlled with regular infusions of the deficient clotting factors. However, clotting factor replacement sometimes leads to the formation of fVIII and fIX inhibitors *in vivo* [71]. Alternatively, recombinant factor VIIa (rFVIIa) has been used to treat bleeding disorders [72, 73] including hemophilia with and without factor VIII/IX inhibitors [74]. However, rFVIIa requires frequent administration (every 2-3 hr), and many questions remain about its mechanism of action, its effective dosage [71], and its overall utility for the treatment trauma-associated hemorrhage [75]. In this study, we did not model rFVIIa-induced coagulation directly. Rather, we modeled a general trigger which initiated the extrinsic coagulation cascade. Since we identified the model using TF/FVIIa, inherent to our rFVIIa simulations (and the rate constant governing initiation) was the presence of TF. However, even with this complication, the model generated potentially useful insight into the rFVIIa mechanism of action, and its possible shortcomings especially for the treatment

of hemophilia. The addition of rFVIIa directly activated thrombin through the initiation pathway. However, no amplification of the thrombin signal occurred without fVIII or fIX. Thus, the peak thrombin signal was lower than normal coagulation, the peak thrombin time was longer, and thrombin generation was eventually inhibited by the combined action of ATIII and the protein C pathway. However, as the dose of rFVIIa increased, the peak thrombin time decreased (eventually saturating around 200×nominal trigger), and the peak thrombin value increased such that the thrombin profile resembled normal coagulation. Butenas *et al.* performed an extensive *in vitro* study of rFVIIa-induced thrombin generation under normal and hemophilic conditions [76]. They found qualitatively similar trends, namely rFVIIa restored normal coagulation (even in the absence of TF) for large enough rFVIIa doses, although rFVIIa-induced coagulation in hemophilia (even for large rFVIIa doses) lagged the normal profile. These results suggest that rFVIIa administration alone might not be able to initiate normal coagulation in recurrent bleeding, unless the dosage is well above a critical threshold. However, defining this threshold, which is likely patient specific, is difficult as there is tremendous patient to patient variability even with a normal co agulation phenotype [77].

A number of mathematical models and computational simulations concerning coagulation dynamics have been based on the Hockin-Mann model of coagulation. Recently, Brummel-Ziedins and co-workers built a model that incorporated the Protein C (PC) pathway into the Hockin-Mann network to investigate thrombin generation in cases of familial PC deficiency [78]. Computational simulations using this model showed that PC mutation caused an elevation in peak thrombin levels without changing the initiation time or the slope of thrombin generation curve. Our hybrid model to a certain extent captured this trend qualitatively. We were able to show the decrease in peak thrombin concentration in the presence of PC pathway without any other inhibitors like ATIII.

Danforth *et al.* computationally simulated normal thrombin generation using 5pM tissue factor with all other factors at their mean physiological level [77]. The initiation time in this simulation was approximately 4.4 minutes. When predicting the normal thrombin generation curve using 0.2 nM of FVIIa we were able to show an approximate initiation time around 5 minutes with a similar thrombin generation profile (Fig. 2.6A). In a kinetic modeling to study the effect of rfVIIa on thrombin generation [79] it was shown that increasing amounts of FVIIa accelerated the initiation phase and the propagation phase during thrombin generation. Analysis of thrombin generation curves showed that the FVIIa controlled the maximal slope of thrombin curve along with peak thrombin time and initiation time. Though we had issues in capturing the effect of FVIIa on initiation time we were able to show that change in concentration of FVIIa affected the maximal thrombin slope as well as propagation phase (Fig. 2.7A).

The performance of the proof of principle coagulation model was impressive given its limited size. The most detailed mechanistic model of coagulation is the one by Luan and co-workers built with 193 proteins and protein complexes that are interconnected by 301 reactions [45]. Akin to the hybrid model this model used normal thrombin data and hemophilia data from Allen *et al.* [3]. However unlike the mechanistic model we used this data for validation rather than training. Results from our model are surprisingly comparable to the training simulations of the mechanistic model. The initiation time and amplification of thrombin signal were accurately predicted in the normal case whereas in the hemophilia case we correctly predict the initiation time but slightly over predict the amplification signal. This performance is nearly similar to that of the detailed mechanistic model with 301 reactions. Despite the impressive performance, there are several issues that could be further explored. First, the prediction of initiation time should be investigated. We were able to estimate initiation time within a data set, but unable to

predict initiation time *across* independent data sets. This suggested that we should update the initiation module to distinguish between different triggers, e.g., TF/FVIIa versus rFVIIa alone, and to include key biological milestones such as FXa activation (a prerequisite to thrombin formation). Next, there are several additional biological modules that could be added to the core model presented here. First, we could include thrombin-induced platelet activation and the role of activated platelets in amplification. We captured thrombin generation data in the presence of platelets, however, the initial shape of the activation curve and the time-scale of activation was not always consistent with the data. Platelets are activated by thrombin through the cleavage of the extracellular domain of protease-activated receptors (PARs) on the platelet surface. Once activated, platelets play an important role in amplification, and are key mediators of the positive feedback driving amplification. Thus, this biology is a potentially important component of an expanded model. We should also add the intrinsic pathway to the model. The intrinsic pathway is triggered by contact activation of the plasma protease factor XI (fXI) by negatively charged surfaces and by thrombin and upstream factors such as activated plasma protease factor XII (FXIIa) [80, 81]. Activated platelets may also release polyphosphate which directly activates fXII [82]. Arguably a minor player in acute bleeding, contact activation could also be important in other wound healing contexts. Finally, to make the model more clinically relevant, we should include the biochemical processes responsible for clot formation and clot dissolution (fibrinolysis). Clot formation is driven by thrombin activity, while fibrinolysis is driven by plasmin activity, a key enzyme that cleaves fibrin (one of the main materials in a clot). Similar to coagulation, fibrinolysis is managed by several activating and inhibitory factors which control the balance between clot formation and dissolution. Tissue plasminogen activator (t-PA) and urokinase activate plasmin, along with contact pathway factors such as fXIa. On the other hand, thrombin activatable fibrinolysis inhibitor (TAFI) inhibits the degradation of fibrin by plasmin. Also, similar

to coagulation, there is considerable fibrinolysis and contact pathway data sets that can be used to train the model. Lastly, the choice of max/min integration rules or the particular form of the transfer functions could be generalized to include other rule types and functions. Theoretically, an integration rule is a function whose domain is a set of transfer function inputs, and whose range is $v \in [0, 1]$. Thus, integration rules other than max/min could be used, such as the mean or the product, assuming the range of the transfer functions is always $f \in [0, 1]$. Alternative integration rules such as the mean might have different properties which could influence model identification or performance. For example, a mean integration rule would be differentiable, which allows derivative-based optimization approaches to be used. The particular form of the transfer function could also be explored. We choose a Hill-like function because of its prominence in the systems and synthetic biology community. However, the only mathematical requirement for a transfer function is that it map a non-negative continuous or categorical variable into the range $f \in [0, 1]$. Thus, many types of transfer functions are possible.

2.4 Materials and Methods

2.4.1 Formulation and Solution of the Model Equations

We used ordinary differential equations (ODEs) to model the time evolution of proteins (x_i) in our reduced order coagulation model:

$$\frac{dx_i}{dt} = \sum_{j=1}^{\mathcal{R}} \sigma_{ij} r_j (\mathbf{x}, \epsilon, \mathbf{k}) \qquad i = 1, 2, \ldots, \mathcal{M} \tag{2.1}$$

where $\mathcal{R}$ denotes the number of reactions, $\mathcal{M}$ denotes the number of protein species in the model. The quantity $r_j(\mathbf{x}, \epsilon, \mathbf{k})$ denotes the rate of reaction j. Typically, reaction j is a non-linear function of biochemical species abundance, as well as unknown kinetic parameters $\mathbf{k}$ ($\mathcal{K} \times 1$). The quantity σ_{ij} denotes the stoichiometric coefficient for species i in reaction j. If $\sigma_{ij} > 0$, species i is produced by reaction j. Conversely, if $\sigma_{ij} < 0$, species i is consumed by reaction j, while $\sigma_{ij} = 0$ indicates species i is not connected with reaction j. The system material balances were subject to the initial conditions $\mathbf{x}(t_o) = \mathbf{x}_o$, which were specified by the experimental setup.

Each reaction rate was written as the product of two terms, a kinetic term ($\bar{r}_j$) and a control term (v_j) that depends on multiple regulators:

$$r_j(\mathbf{x}, \epsilon, \mathbf{k}) = \bar{r}_j v_j \tag{2.2}$$

We used multiple saturation kinetics to model the reaction term $\bar{r}_j$:

$$\bar{r}_j = k_j^{max} \epsilon_i \left(\prod_{s \in m_j^-} \frac{x_s}{K_{js} + x_s} \right) \tag{2.3}$$

where k_j^{max} denotes the maximum rate for reaction j, ϵ_i denotes the scaled enzyme activity which catalyzes reaction j, and K_{js} denotes the saturation constant for species s in reaction j. The product in Eqn. (3.4) was carried out over the set of *reactants* for reaction j (denoted as m_j^-).

The control term v_j depended upon the combination of factors which influenced the activity of enzyme i. For each enzyme, we used a rule-based approach to select from competing control factors (Fig. 3.2). If an enzyme was activated by m metabolites, we

modeled this activation as:

$$v_j = \max\left(f_{1j}(\mathcal{Z}), \ldots, f_{mj}(\mathcal{Z})\right) \tag{2.4}$$

where $0 \leq f_{ij}(\mathcal{Z}) \leq 1$ was a regulatory transfer function that calculated the influence of metabolite i on the activity of enzyme j. Conversely, if enzyme activity was inhibited by m metabolites, we modeled this inhibition as:

$$v_j = 1 - \max\left(f_{1j}(\mathcal{Z}), \ldots, f_{mj}(\mathcal{Z})\right) \tag{2.5}$$

Lastly, if an enzyme had both m activating and n inhibitory factors, we modeled the control term as:

$$v_j = \min\left(u_j, d_j\right) \tag{2.6}$$

where:

$$u_j = \max_{j^+}\left(f_{1j}(\mathcal{Z}), \ldots, f_{mj}(\mathcal{Z})\right) \tag{2.7}$$

$$d_j = 1 - \max_{j^-}\left(f_{1j}(\mathcal{Z}), \ldots, f_{nj}(\mathcal{Z})\right) \tag{2.8}$$

The quantities j^+ and j^- denoted the sets of activating and inhibitory factors for enzyme j. If a process has no modifying factors, we set $v_j = 1$. There are many possible functional forms for $0 \leq f_{ij}(\mathcal{Z}) \leq 1$. However, in this study, each individual transfer function took the form:

$$f_i\left(\mathcal{Z}_j, k_{ij}\right) = \frac{k_{ij}^{\eta}\mathcal{Z}_j^{\eta}}{1 + k_{ij}^{\eta}\mathcal{Z}_j^{\eta}} \tag{2.9}$$

where $\mathcal{Z}_j$ denotes the abundance of the j factor (e.g., metabolite abundance), and k_{ij} and

η are control parameters. k_{ij} was the species gain parameter, while η was a cooperativity parameter (similar to a Hill coefficient). Applying the general framework to the reduced coagulation network resulted in five ordinary differential equations:

$$\frac{dx_1}{dt} = -\left(r_{init}v_{init} + r_{amp}v_{amp}\right) \tag{2.10}$$

$$\frac{dx_2}{dt} = r_{amp}v_{amp} + r_{init}v_{init} - r_{inh,ATIII}v_{inh,ATIII} \tag{2.11}$$

$$\frac{dx_3}{dt} = -r_{apc}v_{apc} \tag{2.12}$$

$$\frac{dx_4}{dt} = r_{apc}v_{apc} \tag{2.13}$$

$$\frac{dx_5}{dt} = -r_{inh,ATIII}v_{inh,ATIII} \tag{2.14}$$

where $\mathbf{x} = (fII, FIIa, PC, APC, ATIII)^T$. The terms r_*v_* in the balance equations denote corrected kinetic expressions for initiation, amplification and inhibition processes. The rate of initiation $\bar{r}_{init}$ was modeled as:

$$\bar{r}_{init} = k_{init}\,(trigger)\,\frac{x_1}{K_{init,fII} + x_1} \tag{2.15}$$

where k_{init}, $K_{init,fII}$ are the rate and saturation constants governing initiation, respectively. The rate of initiation was modified by v_{init}, the control parameter governing initiation. Initiation was sensitive to the level of trigger (activator) and TFPI (inhibitor):

$$v_{init} = \min\left(f^-_{init}\,(TFPI)\,, f^+_{init}\,(trigger)\right) \tag{2.16}$$

where the transfer functions f took the form of Eqn (3.10). The rate of thrombin amplification was given by:

$$\bar{r}_{amp} = k_{amp}\,(x_2)\,\frac{x_1}{K_{amp,fII} + x_1} \tag{2.17}$$

where k_{amp}, $K_{amp,fII}$ denote the rate and saturation constants governing amplification, respectively. The amplification control term, which modified amplification rate, was modeled as a combination of multiple inhibition terms and one activation term:

$$v_{amp} = \min\left(f^{-}_{amp}\left(TFPI\right), f^{-}_{amp}\left(x_4\right), f^{+}_{amp}\left(\mathcal{Z}_{amp}\right)\right) \tag{2.18}$$

where $\mathcal{Z}_{amp} = fV \times fX \times fVIII \times fIX$. Although $f^{+}_{amp}\left(\mathcal{Z}_{amp}\right)$ is an activating term, we included it in the min integration rule; the factors in $\mathcal{Z}_{amp}$ were essential for amplification (if any of these factors was missing the amplification reaction would not occur). Thus, the factors in $\mathcal{Z}_{amp}$ were required components, a classification that we implemented by the min selection rule. The rate activated protein C formation was given by:

$$\bar{r}_{apc} = k_{APC,formation}\left(TM\right)\frac{x_3}{K_{formation,PC} + x_3} \tag{2.19}$$

where $k_{APC,formation}$ and $K_{formation,PC}$ denote the rate and saturation constants governing activated protein C formation, respectively and TM denotes the thrombomodulin abundance. We modeled the control term which governed APC formation as a single thrombin-dependent activation term:

$$v_{apc} = \max\left(f^{+}_{apc}\left(x_2\right)\right) \tag{2.20}$$

Lastly, we included direct irreversible inhibition of FIIa by ATIII:

$$\bar{r}_{inh,ATIII} = k_{ATIII,inhibition}\left(x_5 x_2^{\gamma}\right) \tag{2.21}$$

where γ was estimated to be $\gamma = 1.26$. For ATIII inhibition of FIIa, the control variables $v_{inh,ATIII}$ was taken to be unity. The model equations were encoded using the Python pro-

gramming language and solved using the ODEINT routine of the SciPy module [83]. The model files can be downloaded from http://www.varnerlab.org.

2.4.2 Estimation of Model Parameters From Experimental Data

Model parameters were estimated by minimizing the difference between simulations and experimental thrombin measurements (squared residual):

$$\min_{\mathbf{k}} \sum_{\tau=1}^{\mathcal{T}} \sum_{j=1}^{\mathcal{S}} \left(\frac{\hat{x}_j(\tau) - x_j(\tau, \mathbf{k})}{\omega_j(\tau)} \right)^2 \tag{2.22}$$

where $\hat{x}_j(\tau)$ denotes the measured value of species j at time τ, $x_j(\tau, \mathbf{k})$ denotes the simulated value for species j at time τ, and $\omega_j(\tau)$ denotes the experimental measurement variance for species j at time τ. The outer summation is with respect to time, while the inner summation is with respect to state. We minimized the model residual using Particle swarm optimization (PSO) [22]. PSO uses a *swarming* metaheuristic to explore parameter spaces. A strength of PSO is its ability to find the global minimum, even in the presence of potentially many local minima, by communicating the local error landscape experienced by each particle collectively to the swarm. Thus, PSO acts both as a local and a global search algorithm. For each iteration, particles in the swarm compute their local error by evaluating the model equations using their specific parameter vector realization. From each of these local points, a globally best error is identified. Both the local and global error are then used to update the parameter estimates of each particle using the rules:

$$\mathbf{\Delta}_i = \theta_1 \mathbf{\Delta}_i + \theta_2 \mathbf{r}_1 (\mathcal{L}_i - \mathbf{k}_i) + \theta_3 \mathbf{r}_2 (\mathcal{G} - \mathbf{k}_i) \tag{2.23}$$

$$\mathbf{k}_i = \mathbf{k}_i + \mathbf{\Delta}_i \tag{2.24}$$

where $(\theta_1, \theta_2, \theta_3)$ are adjustable parameters, $\mathcal{L}_i$ denotes the local best solution found by particle i, and $\mathcal{G}$ denotes the best solution found over the entire population of particles. The quantities r_1 and r_2 denote uniform random vectors with the same dimension as the number of unknown model parameters ($\mathcal{K} \times 1$). In thus study, we used $(\theta_1, \theta_2, \theta_3) = (1.0, 0.05564, 0.02886)$. The quality of parameter estimates was measured using goodness of fit (model residual). The particle swarm optimization routine was implemented in the Python programming language. All plots were made using the Matplotlib module of Python [84].

2.4.3 Global Sensitivity Analysis of Model Performance

We conducted a global sensitivity analysis, using the variance-based method of Sobol, to estimate which parameters controlled the performance of the reduced order model [85]. We computed the total sensitivity index of each parameter relative to two performance objectives, the peak thrombin time and the area under the thrombin curve (thrombin exposure). We established the sampling bounds for each parameter from the minimum and maximum value of that parameter in the parameter set ensemble. We used the sampling method of Saltelli *et al.* [86] to compute a family of $N\,(2d + 2)$ parameter sets which obeyed our parameter ranges, where N was the number of trials, and d was the number of parameters in the model. In our case, N = 10,000 and d = 22, so the total sensitivity indices were computed from 460,000 model evaluations. The variance-based sensitivity analysis was conducted using the SALib module encoded in the Python programming language [87].

CHAPTER 3

DYNAMIC MODELING OF CELL-FREE BIOCHEMICAL NETWORKS USING EFFECTIVE KINETIC MODELS

Abstract

Cell-free systems [1] offer many advantages for the study, manipulation and modeling of metabolism compared to *in vivo* processes. Many of the challenges confronting genome-scale kinetic modeling can potentially be overcome in a cell-free system. For example, there is no complex transcriptional regulation to consider, transient metabolic measurements are easier to obtain, and we no longer have to consider cell growth. Thus, cell-free operation holds several significant advantages for model development, identification and validation. Theoretically, genome-scale cell-free kinetic models may be possible for industrially important organisms, such as *E. coli*, if a simple, tractable framework for integrating allosteric regulation with enzyme kinetics can be formulated. Toward this unmet need, we present an effective biochemical network modeling framework for building dynamic cell-free metabolic models. The key innovation of our approach is the integration of simple effective rules encoding complex allosteric regulation with traditional kinetic pathway modeling. We tested our approach by modeling the time evolution of several hypothetical cell-free metabolic networks. We found that simple effective rules, when integrated with traditional enzyme kinetic expressions, captured complex allosteric patterns such as ultrasensitivity or non-competitive inhibition in the absence of mechanistic information. Second, when integrated into network models, these rules captured classic regulatory patterns such as product-induced feedback inhibition. Lastly, we showed, at

[1] Adapted with permission from Wayman JA, Sagar A, and Varner JD, "Dynamic modeling of cell-free biochemical networks using effective kinetic models" (2015) Processes, **3**:138-160

least for the network architectures considered here, that we could simultaneously estimate kinetic parameters and allosteric connectivity from synthetic data starting from an unbiased collection of possible allosteric structures using particle swarm optimization. However, when starting with an initial population that was heavily enriched with incorrect structures, our particle swarm approach could converge to an incorrect structure. While only an initial proof-of-concept, the framework presented here could be an important first step toward genome-scale cell-free kinetic modeling of the biosynthetic capacity of industrially important organisms.

3.1 Introduction

Mathematical modeling has long contributed to our understanding of metabolism. Decades before the genomics revolution, mechanistically, structured metabolic models arose from the desire to predict microbial phenotypes resulting from changes in intracellular or extracellular states [88]. The single cell *E. coli* models of Shuler and coworkers pioneered the construction of large-scale, dynamic metabolic models that incorporated multiple, regulated catabolic and anabolic pathways constrained by experimentally determined kinetic parameters [89]. Shuler and coworkers generated many single cell kinetic models, including single cell models of eukaryotes [90, 91], minimal cell architectures [92], as well as DNA sequence based whole-cell models of *E. coli* [93]. Conversely, highly abstracted kinetic frameworks, such as the cybernetic framework, represented a paradigm shift, viewing cells as growth-optimizing strategists [94]. Cybernetic models have been highly successful at predicting metabolic choice behavior, e.g., diauxie behavior [95], steady-state multiplicity [96], as well as the cellular response to metabolic engineering modifications [97]. Unfortunately, traditional, fully structured cybernetic models

also suffer from an identifiability challenge, as both the kinetic parameters and an abstracted model of cellular objectives must be estimated simultaneously. However, recent cybernetic formulations from Ramkrishna and colleagues have successfully treated this identifiability challenge through elementary mode reduction, though the techniques replace detailed biological mechanism with an optimization heuristic [98, 99].

In the post genomics world, large-scale stoichiometric reconstructions of microbial metabolism popularized by static, constraint-based modeling techniques such as flux balance analysis (FBA) have become standard tools [100]. Since the first genome-scale stoichiometric model of *E. coli*, developed by Edwards and Palsson [101], well over 100 organisms, including industrially important prokaryotes such as *E. coli* [102] or *B. subtilis* [103], are now available [104]. Stoichiometric models rely on a pseudo-steady-state assumption to reduce unidentifiable genome-scale kinetic models to an underdetermined linear algebraic system, which can be solved efficiently even for large systems. Traditionally, stoichiometric models have also neglected explicit descriptions of metabolic regulation and control mechanisms, instead opting to describe the choice of pathways by prescribing an objective function on metabolism. Interestingly, similar to early cybernetic models, the most common metabolic objective function has been the optimization of biomass formation [105], although other metabolic objectives have also been estimated [106]. Recent advances in constraint-based modeling have overcome the early shortcomings of the platform, including capturing metabolic regulation and control [107]. Thus, modern constraint-based approaches have proven extremely useful in the discovery of metabolic engineering strategies and represent the state of the art in metabolic modeling [108, 109]. However, genome-scale kinetic models of industrial important organisms such as *E. coli* have yet to be constructed.

Cell-free systems offer many advantages for the study, manipulation and modeling of

metabolism compared to *in vivo* processes. Central amongst these advantages is direct access to metabolites and the microbial biosynthetic machinery without the interference of a cell wall. This allows us to control as well as interrogate the chemical environment while the biosynthetic machinery is operating, potentially at a fine time resolution. Second, cell-free systems also allow us to study biological processes without the complications associated with cell growth. Cell-free protein synthesis (CFPS) systems are arguably the most prominent examples of cell-free systems used today [110]. However, CFPS is not new; CFPS in crude *E. coli* extracts has been used since the 1960s to explore fundamentally important biological mechanisms [111, 112]. Today, cell-free systems are used in a variety of applications ranging from therapeutic protein production [113] to synthetic biology [114]. Interestingly, many of the challenges confronting genome-scale kinetic modeling can potentially be overcome in a cell-free system. For example, there is no complex transcriptional regulation to consider, transient metabolic measurements are easier to obtain, and we no longer have to consider cell growth. Thus, cell-free operation holds several significant advantages for model development, identification and validation. Theoretically, genome-scale cell-free kinetic models may be possible for industrially important organisms, such as *E. coli* or *B. subtilis*, if a simple, tractable framework for integrating allosteric regulation with enzyme kinetics can be formulated.

In this study, we present an effective biochemical network modeling framework for building dynamic cell-free metabolic models. The key innovation of our approach is the seamless integration of simple effective rules encoding complex regulation with traditional kinetic pathway modeling. This integration allows the description of complex regulatory interactions, such as time-dependent allosteric regulation of enzyme activity, in the absence of specific mechanistic information. The regulatory rules are easy to understand, easy to formulate and do not rely on overarching theoretical abstractions or

restrictive assumptions. We tested our approach by modeling the time evolution of several hypothetical cell-free metabolic networks. In particular, we tested whether our effective modeling approach could describe classically expected enzyme kinetic behavior, and second whether we could simultaneously estimate kinetic parameters and regulatory connectivity, in the absence of specific mechanistic knowledge, from synthetic experimental data. Toward these questions, we explored five hypothetical cell-free networks. Each network shared the same enzymatic connectivity, but had different allosteric regulatory connectivity. We found that simple effective rules, when integrated with traditional enzyme kinetic expressions, captured complex allosteric patterns such as ultrasensitivity or non-competitive inhibition in the absence of mechanistic information. Second, when integrated into network models, these rules captured classical regulatory patterns such as product-induced feedback inhibition. Lastly, we showed, at least for the network architectures considered here, that we could simultaneously estimate kinetic parameters and allosteric connectivity from synthetic data starting from an unbiased collection of possible allosteric structures using particle swarm optimization. However, when starting with an initial population that was heavily enriched with incorrect structures, our particle swarm approach could converge to an incorrect structure. While only an initial proof-of-concept, the framework presented here could be an important first step toward genome-scale cell-free kinetic modeling of the biosynthetic capacity of industrially important organisms.

3.2 Results

3.2.1 Formulation and properties of effective cell-free metabolic models.

We developed two proof-of-concept metabolic networks to investigate the features of our effective biochemical network modeling approach (Fig. 3.1). In both examples, substrate S was converted to the end products P_1 and P_2 through a series of enzymatically catalyzed reactions, including a branch point at hypothetical metabolite M_2. Several of these reactions involved cofactor dependence (AH or A), and various allosteric regulatory mechanisms modified the activity of pathway enzymes. Network A included feedback inhibition of the initial pathway enzyme (E_1) by pathway end products P_1 and P_2 (Fig. 3.1A). On the other hand, network B involved feedback inhibition of E_1 by P_2 and E_6 by P_1 (Fig. 3.1B). In both networks, branch point enzymes E_3 and E_6 were subject to feed-forward activation by reduced cofactor AH. Lastly, it is known experimentally that cell-free systems have a finite operational lifespan. Loss of biosynthetic capability could be a function of many factors, e.g., cofactor or metabolite limitations. We modeled the loss of biosynthetic capability as a non-specific first-order decay of enzyme activity.

Allosteric regulation of enzyme activity was modeled by combining individual regulatory contributions to the activity of pathway enzymes into a control coefficient using an integration rule (Fig. 3.2). This strategy is similar in spirit to the Constrained Fuzzy Logic (cFL) approach of Lauffenburger and coworkers which has been used to effectively model signal transduction pathways important in human health [115]. In our formulation, Hill-like transfer functions $0 \leq f(\mathcal{Z}) \leq 1$ were used to calculate the influence of factor abundance upon target enzyme activity. In this context, factors can be individual metabolite

levels or some function, e.g., the product of metabolite levels. However, more generally, factors can also correspond to non-modeled influences, categorial variables or other abstract quantities. In the current study, we simply let $\mathcal{Z}$ correspond to the abundance of individual metabolites, however in general this can be a complex function of both modeled and unmodeled factors. When an enzyme was potentially sensitive to more than one regulatory input, logical integration rules were used to select which regulatory transfer function influenced enzyme activity at any given time. Thus, our test networks involved important features such as cofactor recycling, enzyme activity and metabolite dynamics, as well as multiple overlapping allosteric regulatory mechanisms.

The rule-based regulatory strategy approximated the behavior of classical allosteric activation and inhibition mechanisms (Fig. 3.3). We considered the enzyme catalyzed conversion of substrate S to a product P, where the overall reaction rate was modeled as the product of a Michaelis-Menten term and an effective allosteric control variable reflecting the particular regulatory interaction. We first explored feed-forward substrate activation of enzyme activity (for both positive and negative cooperativity). Consistent with classical data, the rule-based strategy predicted a sigmoidal relationship between substrate abundance and reaction rate as a function of the cooperativity parameter (Fig. 3.3A). For cooperativity parameters less than unity, increased substrate abundance *decreased* the maximum reaction rate. This was consistent with the idea that substrate binding *decreased* at regulatory sites, which negatively impacted substrate binding at the active site. On the other hand, as the cooperativity parameter increased past unity, the rate of conversion of substrate S to product P by enzyme E approached a step function. In the presence of an inhibitor, the rule-based strategy predicted non-competitive like behavior as a function of the cooperativity parameter (Fig. 3.3B). When the control gain parameter, κ_{ij} in Eqn. (3.10), was greater than unity, the inhibitory force was directly

proportional to the cooperativity parameter, η in Eqn. (3.10). Thus, as the cooperativity parameter increased, the maximum reaction rate decreased (Fig. 3.3B). Interestingly, our rule-based approach was unable to directly simulate competitive inhibition of enzyme activity. Taken together, the rule-based strategy captured classical regulatory patterns for both enzyme activation and inhibition. Thus, we are able to model complex kinetic phenomena such as ultrasensitivity, despite an effective description of reaction kinetics.

End product yield was controlled by feedback inhibition, while product selectivity was controlled by branch point enzyme inhibition (Fig. 3.4). A critical test of our modeling approach was to simulate networks with known behavior. If we cannot reproduce the expected behavior of simple networks, then our effective modeling strategy, and particularly the rule-based approximation of allosteric regulation, will not be feasible for genome-scale cell-free problems. We considered two cases, control ON/OFF, for each network configuration. Each of these cases had identical kinetic parameters and initial conditions; the *only* differences between the cases were the allosteric regulation rules and the control parameters associated with these rules. As expected, end product accumulation was larger for network A when the control was OFF (no feedback inhibition of E_1 by P_1 and P_2), as compared to the ON case (Fig. 3.4A). We found this behavior was robust to the choice of underlying kinetic parameters, as we observed that same qualitative response across an ensemble of 100 randomized parameter sets, for fixed control parameters. The control ON/OFF response of network B was more subtle. In the OFF case, the behavior was qualitatively similar to network A. However, for the ON case, flux was diverted away from P_2 formation by feedback inhibition of E_6 activity at the M_2 branch point by P_1 (Fig. 3.4B). Lower E_6 activity at the M_2 branch point allowed more flux toward P_1 formation, hence the yield of P_1 also increased (Fig. 3.4C). Again, the control ON/OFF behavior of network B was robust to changes in kinetic parameters, as the same quali-

tative trend was conserved across an ensemble of 100 randomized parameters, for fixed control parameters. Taken together, these simulations suggested that the rule-based allosteric control concept could robustly capture expected feedback behavior for networks with uncertain kinetic parameters.

3.2.2 Estimating parameters and effective allosteric regulatory structures.

A critical challenge for any dynamic model is the estimation of kinetic parameters. For metabolic processes, there is also the added challenge of identifying the regulation and control structures that manage metabolism. Of course, these issues are not independent; any description of enzyme activity regulation will be a function of system state, which in turn depends upon the kinetic parameters. For cell-free systems, regulated gene expression has been removed, however, enzyme activity regulation is still operational. We explored this linkage by estimating model parameters from synthetic data using both network structures. We generated synthetic measurements of the substrate S, intermediate M_5 and end product P_1 approximately every 20 min using network A. This data set is similar to published cell-free studies, both in terms of network coverage and sampling frequency [110]. We then generated an ensemble of model parameter estimates by minimizing the difference between model simulations and the synthetic data using particle swarm optimization (PSO), starting from random initial parameter guesses. The estimation of kinetic parameters was sensitive to the choice of regulatory structure (Fig. 3.5). PSO identified an ensemble of parameters that bracketed the mean of the synthetic measurements in less than 1000 iterations when the control structure was correct (Fig. 3.5A and B). However, with control mismatch (network B simulated with network A param-

eters), model simulations were not consistent with the synthetic data (Fig. 3.5C and D). Taken together, these results suggested that we could perhaps simultaneously estimate both parameters and network control architectures, as incorrect control structures would be manifest as poor model fits.

We modified our particle swarm identification strategy to simultaneously search over both kinetic parameters and putative control structures. In addition to our initial networks, we constructed three additional presumptive network models, each with the same enzymatic connectivity but different allosteric regulation of the pathway enzymes (Fig. 3.6). We then initialized a population of particles, each with one of the five potential regulatory programs and randomized kinetic parameters. Thus, we generated an initial population of particles that had *both* different kinetic parameters as well as different control structures. We biased the distribution of the particle population according to our *a prior* belief of the correct regulatory program. To this end, we considered three different priors, a uniform distribution where each putative regulatory structure represented 20% of the population and two mixed distributions that were either positively or negatively biased towards the correct structure (network A). In both the positively biased and uniform cases the PSO clearly differentiated between the true or closely related structures and those that were materially different (Fig. 3.7). As expected, the positively biased population (40% of the initial particle population seeded with network A) gave the best results, where the correct structure was preferentially identified (Fig. 3.7A). On the other hand, when given a uniform distribution, the PSO approach identified a combination of network A and network C as the most likely control structures (Fig. 3.7B). Network A and C differ by the regulatory connection between the end product P_2 and enzyme E_1; in network A, end product P_2 was assumed to inhibit E_1, while in network C, end product P_2 activated E_1. Lastly, when the initial population was heavily biased towards incorrect

structures (initial population seeded with 90% incorrect structures), the particle swarm *misidentified* the correct allosteric structure (Fig. 3.7C). Interestingly, while each particle swarm identified parameter sets that minimized the simulation error, the estimated parameter values were not necessarily similar to the true parameters. The angle between the estimated and true parameters was not consistently small across the swarms (identical parameters would give an angle of zero). This suggested that our particle swarm approach identified a *sloppy* ensemble, i.e., parameter estimates that were individually incorrect but collectively exhibited the correct model behavior.

We calculated control program output and scaled metabolic flux for the positively, uniformly and negatively biased particle swarms (Fig. 3.8). Network A and network C models from the positively (Fig. 3.8A) and uniformly (Fig. 3.8B) biased particle swarms showed similar operational patterns, despite differences in kinetic parameters and control structures. While models from the negatively biased population had error values similar to the correct structures in the previous swarms, they have different flux and control profiles (Fig. 3.8C). In all cases, regardless of network configuration or parameter values, the rate of enzyme decay was small compared to the other fluxes, and all networks had qualitatively similar trends for E_3 and E_6 control. Moreover, consistent with the correct model structure, production of end product P_1 was the preferred branch for all model configurations. However, there was variability in P_2 production flux across the population of models, especially for the uniform swarm when compared with the other cases. High P_1 branch flux resulted in end product inhibition of E_1 in both network A and network C, however in network D and E, high P_1 flux induced E_1 activation. These trends were manifested in different flux profiles, where the negatively biased population appeared more uniform across the population compared with the other swarms, and had higher E_1 specific activity. Interestingly, the behavior of network A and network C highlighted an

artifact of our integration rule; both a positive or negative feedback connection from P_2 to E_1 were ignored because the P_1 inhibition of E_1 dominated. Thus, while theoretically distinct, network A and network C appeared operationally to the PSO algorithm to be the same network. On the other hand, networks B, D and E showed distinct behavior that was not consistent with the true network. These architectures exhibited either limited inhibition (network B) or activation (network D and E) of E_1 activity, resulting in significantly different metabolic flux profiles. However, the PSO was able to find low error parameter solutions, despite the mismatch in the control structures (error values similar, but not better than the best network A and network C estimates). Taken together, these results suggested that a uniform sampling approach could potentially yield an unbiassed estimate of both kinetic parameters and control structures. However, the negatively biased particle swarm results illustrated a potential shortcoming of the approach, namely convergence to a local error minimum despite a significantly incorrect control structure. This suggested that estimated model structures will need to be further evaluated, for example by generating falsifiable experimental designs which could distinguish between low error solutions.

3.3 Discussion

In this study, we presented an effective kinetic modeling strategy to dynamically simulate cell-free biochemical networks. Our proposed strategy integrated traditional kinetic modeling with an effective rules based approach to dynamically describe metabolic regulation and control. We tested this approach by developing kinetic models of hypothetical cell-free metabolic networks. In particular, we tested whether our effective modeling approach could describe classically expected behavior, and second whether we could simul-

taneously estimate kinetic parameters and regulatory connectivity, in the absence of specific mechanistic knowledge, from synthetic experimental data. Toward these questions, we explored five hypothetical cell-free networks. In each network, a substrate S was converted to the end products P_1 and P_2 through a series of enzymatically catalyzed reactions, including a branch point at a hypothetical metabolite M_2. Each network also included the same cofactors and cofactor recycle architecture. However, while all five networks shared the same enzymatic connectivity, each had different allosteric regulatory connectivity. We found that simple effective rules, when integrated with traditional enzyme kinetic expressions, could capture complex allosteric patterns such as ultrasensitivity, or non-competitive inhibition in the absence of specific mechanistic information. Moreover, when integrated into network models, these rules captured classical regulatory patterns such as product-induced feedback inhibition. Lastly, we simultaneously estimated kinetic parameters and discriminated between competing regulatory structures, using synthetic data in combination with a modified particle swarm approach. If we considered all putative regulatory architectures to be equally likely, we were able to estimate a *sloppy* ensemble of models with the correct architecture and kinetic parameters. Thus, we identified parameter values that were different from their true values, but nonetheless produced reasonable model performance (low error). This suggested that we captured important parameter combinations (stiff combinations), while simultaneously missing other parameter combinations (sloppy combinations). This was similar to the earlier study of Brown and Sethna [116], which showed that reasonable model predictions were possible, despite sometimes only order of magnitude parameter estimates, if the stiff parameter combinations were well constrained.

The proposed modeling strategy shares features with other popular techniques, but also has several key differences. At its core, our effective modeling approach is sim-

ilar to regulatory constraint-based methods, and to the cybernetic modeling paradigm developed by Ramkrishna and colleagues. Covert, Palsson and coworkers drastically improved the predictability of constraint-based approaches by integrating Boolean rules into the calculation of metabolic fluxes [117]. If the regulated intracellular flux problem is coupled with time-dependent extracellular balances, these models can predict complex behavior such as diauxie growth or the switch between aerobic and anaerobic metabolism. Another important feature of this approach is that it scales with biological complexity. For example, Covert *et al.* showed that a genome-scale model of *E. coli* augmented with a Boolean rule layer, correctly predicted approximately 80% of the outcomes of a high-throughput growth phenotyping experiment in *E. coli*. Further, they showed that they could learn new biology by iteratively refining the model and its associated rules [118]. However, while regulated flux balance analysis is a powerful technique, it does not easily allow the calculation of time-resolved metabolite abundance. Additionally, the Boolean rules which populate the regulatory layer are limited to ON/OFF decisions; for qualitative predictions of gene expression this is a reasonable limitation. However, Boolean rules will likely be less effective at capturing dynamic allosteric regulation in a cell-free metabolic system. On the other hand, the strength of cybernetic models is the integration of optimal metabolic control heuristics with traditional kinetic pathway modeling. Cybernetic models are highly predictive; they have successfully predicted mutant behavior from limited wild-type data [97, 119, 120], steady-state multiplicity [96], strain specific metabolic function [99] and have been used in bioprocess control applications [121]. However, cybernetic control heuristics are not mechanistic, instead they are the output of an optimal decision with respect to a set of hypothetical physiological objectives. Thus, they are abstractions which are difficult to translate into a specific biological mechanism. Our approach addresses the shortcomings of both regulatory constraint-based models and cybernetic models. First, similar to cybernetic models, the core of our approach is a kinetic

model. Thus, we are able to directly calculate the time evolution of metabolism, for example the dynamic abundance of network metabolites. Second, similar to regulatory flux balance analysis, our control laws describe specific mechanistic motifs, such as activation or inhibition of enzyme activity. However, our rules are continuous, thus they potentially allow a finer grained description of metabolic regulation and control mechanisms. Lastly, we can naturally incorporate unmodeled factors and categorical factors or combinations thereof into our control law formulations. Though requiring a more complex description of cellular metabolism, our approach may even be extended to simulate cell-based systems by incorporating the same control laws into transcription factor activation and gene expression regulation.

There are several critical questions that should be explored following this proof-of-concept study. It is unclear how parameter identification will scale to genome-scale networks, and second it is unclear how we will identify allosteric connectivity at a genome-scale. The enzymatic connectivity for genome-scale cell-free networks can easily be established by stripping away the growth and cell wall machinery from whole cell genome reconstructions. Then metabolic fluxes can be transformed into kinetic expressions using heuristics such multiple saturation kinetics, which are then modified by our rule-based control variables. This leaves a large number of unknown kinetic constants that must be estimated from time-resolved metabolite measurements. Ensemble modeling is a well-established approach for parameter identification in large-scale deterministic models. Liao and coworkers developed a method that generates an ensemble of kinetic models that all approach the same steady-state, one determined by fluxomics measurements [65]. The best subpopulation of candidate models were selected based on their agreement with further measurements of genetically perturbed systems. Our work relies on heuristic search optimization to identify kinetic models consistent with steady-state and dynamic

time-series measurements of cellular species [26, 44, 61, 63, 64, 122]. Instead of estimating a single yet highly uncertain parameter set, both approaches estimate an ensemble of parameter sets whose model behavior recapitulates experimental measurements. Here, we showed that particle swarm optimization quickly identified an ensemble of model parameters, at least for proof-of-concept metabolic networks using synthetic data. This suggested that we can expect reasonable model predictions, despite only partial parameter knowledge, as network size grows if we have properly designed experiments. Brown and Sethna showed in a model of signal transduction that good predictions were possible despite only order of magnitude estimates of parameter values [116]. Sethna and coworkers later showed that model performance is often controlled by only a few parameter combinations, a characteristic seemingly universal to multi-parameter models referred to as *sloppiness* [25]. We have also demonstrated *sloppy* behavior in a wide variety of signal transduction processes [26, 44, 61, 63, 64, 122]. Thus, given our previous experience with models containing hundreds of unknown parameters, we expect parameter estimation to be a manageable challenge assuming we have good quality experimental data.

A second critical challenge will be the estimation of allosteric connectivity at a genome scale. The regulation of glycolytic enzymes, such as phosphofructokinase I, has been studied for many years [123, 124]. The allosteric regulation of metabolic enzymes can also be established from organism specific databases, such as EcoCyc [125], or more general allosteric databases, such as the AlloSteric Database [126]. However, for those enzymes that have not been well studied, we will need to infer allosteric interactions from experimental data. In general, the reverse engineering of regulatory network structure from data is a difficult problem. Recently, Sauer and colleagues have developed a systematic, model-based approach for the identification of allosteric regulation *in vivo* [127]. They tested the effects of many putative allosteric protein-metabolite interactions on the

performance of a kinetic model of glycolysis against dynamic metabolomic and fluxomic measurements. A method similar to this may be easily applied to cell-free systems in order to identify relevant *in vitro* allosteric interactions. Because omics measurements of cell-free environments are easy to obtain, identification of large-scale allosteric control structures may be possible. Also, there are many different approaches from the reverse engineering of gene regulatory networks that perhaps could be adopted to this problem, however this remains an open question. For example, one could imagine designing pulse chase experiments which maximally distinguish between competing allosteric models, similar to the earlier work of Kremling et al [128], or iteratively estimating model structures similar to Doyle and coworkers [129]. Lastly, the choice of max/min integration rules or the particular form of the transfer functions could be generalized to include other rule types and functions. Theoretically, an integration rule is a function whose domain is a set of transfer function inputs, and whose range is $v \in [0, 1]$. Thus, integration rules other than max/min could be used, such as the mean or the product, assuming the range of the transfer functions is always $f \in [0, 1]$. Alternative integration rules such as the mean might have different properties which could influence model identification or performance. For example, a mean integration rule would be differentiable, which allows derivative-based optimization approaches to be used. The particular form of the transfer function could also be explored. We choose a Hill-like function because of its prominence in the systems and synthetic biology community. However, the only mathematical requirement for a transfer function is that it map a non-negative continuous or categorical variable into the range $f \in [0, 1]$. Thus, many types of transfer functions are possible.

3.4 Materials and Methods

3.4.1 Formulation and solution of the model equations.

We used ordinary differential equations (ODEs) to model the time evolution of metabolite (x_i) and scaled enzyme abundance (ϵ_i) in hypothetical cell-free metabolic networks:

$$\frac{dx_i}{dt} = \sum_{j=1}^{\mathcal{R}} \sigma_{ij} r_j(\mathbf{x}, \epsilon, \mathbf{k}) \qquad i = 1, 2, \ldots, \mathcal{M} \tag{3.1}$$

$$\frac{d\epsilon_i}{dt} = -\lambda_i \epsilon_i \qquad i = 1, 2, \ldots, \mathcal{E} \tag{3.2}$$

where $\mathcal{R}$ denotes the number of reactions, $\mathcal{M}$ denotes the number of metabolites and $\mathcal{E}$ denotes the number of enzymes in the model. The quantity $r_j(\mathbf{x}, \epsilon, \mathbf{k})$ denotes the rate of reaction j. Typically, reaction j is a non-linear function of metabolite and enzyme abundance, as well as unknown kinetic parameters $\mathbf{k}$ ($\mathcal{K} \times 1$). The quantity σ_{ij} denotes the stoichiometric coefficient for species i in reaction j. If $\sigma_{ij} > 0$, metabolite i is produced by reaction j. Conversely, if $\sigma_{ij} < 0$, metabolite i is consumed by reaction j, while $\sigma_{ij} = 0$ indicates metabolite i is not connected with reaction j. Lastly, λ_i denotes the scaled enzyme degradation constant. The system material balances were subject to the initial conditions $\mathbf{x}(t_o) = \mathbf{x}_o$ and $\epsilon(t_o) = \mathbf{1}$ (initially we have 100% cell-free enzyme abundance).

Each reaction rate was written as the product of two terms, a kinetic term ($\bar{r}_j$) and a regulatory term (v_j):

$$r_j(\mathbf{x}, \epsilon, \mathbf{k}) = \bar{r}_j v_j \tag{3.3}$$

We used multiple saturation kinetics to model the reaction term $\bar{r}_j$:

$$\bar{r}_j = k_j^{max} \epsilon_i \left(\prod_{s \in m_j^-} \frac{x_s}{K_{js} + x_s} \right) \tag{3.4}$$

where k_j^{max} denotes the maximum rate for reaction j, ϵ_i denotes the scaled enzyme activity which catalyzes reaction j, and K_{js} denotes the saturation constant for species s in reaction j. The product in Eqn. (3.4) was carried out over the set of *reactants* for reaction j (denoted as m_j^-).

The allosteric regulation term v_j depended upon the combination of factors which influenced the activity of enzyme i. For each enzyme, we used a rule-based approach to select from competing control factors (Fig. 3.2). If an enzyme was activated by m metabolites, we modeled this activation as:

$$v_j = \max \left(f_{1j}(\mathcal{Z}), \ldots, f_{mj}(\mathcal{Z}) \right) \tag{3.5}$$

where $0 \leq f_{ij}(\mathcal{Z}) \leq 1$ was a regulatory transfer function that calculated the influence of metabolite i on the activity of enzyme j. Conversely, if enzyme activity was inhibited by a m metabolites, we modeling this inhibition as:

$$v_j = 1 - \max \left(f_{1j}(\mathcal{Z}), \ldots, f_{mj}(\mathcal{Z}) \right) \tag{3.6}$$

Lastly, if an enzyme had both m activating and n inhibitory factors, we modeled the regulatory term as:

$$v_j = \min \left(u_j, d_j \right) \tag{3.7}$$

where:

$$u_j = \max_{j^+} \left(f_{1j}(\mathcal{Z}), \ldots, f_{mj}(\mathcal{Z}) \right) \tag{3.8}$$

$$d_j = 1 - \max_{j^-} \left(f_{1j}(\mathcal{Z}), \ldots, f_{nj}(\mathcal{Z}) \right) \tag{3.9}$$

The quantities j^+ and j^- denoted the sets of activating and inhibitory factors for enzyme j. If an enzyme had no allosteric factors, we set $v_j = 1$. There are many possible functional forms for $0 \leq f_{ij}(\mathcal{Z}) \leq 1$. However, in this study, each individual transfer function took the form:

$$f_i(\mathbf{x}) = \frac{\kappa_{ij}^{\eta} \mathcal{Z}_j^{\eta}}{1 + \kappa_{ij}^{\eta} \mathcal{Z}_j^{\eta}} \tag{3.10}$$

where $\mathcal{Z}_j$ denotes the abundance of the j factor (e.g., metabolite abundance), and κ_{ij} and η are control parameters. The κ_{ij} parameter represents a species gain parameter, while η is a cooperativity parameter (similar to a Hill coefficient). In the case $\eta > 1$, the allosteric interaction displays positive cooperativity. For $\eta < 1$, the interaction is negatively cooperative. Finally, if $\eta = 1$, the interaction displays no cooperativity. The effect of different values of η on reaction rate can be seen in Figure 3.3. The model equations were encoded using the Octave programming language and solved using the LSODE routine in Octave (v 3.8.1; www.octave.org). In some cases, metabolic fluxes (or other quantities) were scaled according to:

$$\hat{r}_j(t = \tau) = \left(\frac{r_j - \min \mathbf{r}}{\max \mathbf{r} - \min \mathbf{r}} \right) \Bigg|_{t=\tau} \tag{3.11}$$

where $0 \leq \hat{r}_j(t = \tau) \leq 1$ denotes the scaled value for flux j evaluated at time τ. We have used this scaling in a variety of other contexts [64, 130].

Estimation of model parameters and structures from synthetic experimental data.

Model parameters were estimated by minimizing the difference between simulations and synthetic experimental data (squared residual):

$$\min_{\mathbf{k}} \sum_{\tau=1}^{\mathcal{T}} \sum_{j=1}^{\mathcal{S}} \left(\frac{\hat{x}_j(\tau) - x_j(\tau, \mathbf{k})}{\omega_j(\tau)} \right)^2 \tag{3.12}$$

where $\hat{x}_j(\tau)$ denotes the measured value of species j at time τ, $x_j(\tau, \mathbf{k})$ denotes the simulated value for species j at time τ, and $\omega_j(\tau)$ denotes the experimental measurement variance for species j at time τ. The outer summation is respect to time, while the inner summation is with respect to state. We approximated a realistic model identification scenario, assuming noisy experimental data, limited sampling resolution (approximately 20 minutes per sample) and a limited number of measurable metabolites. We assumed a constant coefficient of variation of 10% for the synthetic data set.

We minimized the model residual using particle swarm optimization (PSO) [22]. PSO uses a *swarming* metaheuristic to explore parameter spaces. A strength of PSO is its ability to find the global minimum, even in the presence of potentially many local minima, by communicating the local error landscape experienced by each particle collectively to the swarm. Thus, PSO acts both as a local and a global search algorithm. For each iteration, particles in the swarm compute their local error by evaluating the model equations using their specific parameter vector realization. From each of these local points, a globally best error is identified. Both the local and global error are then used to update the parameter

estimates of each particle using the rules:

$$\boldsymbol{\Delta}_i = \theta_1 \boldsymbol{\Delta}_i + \theta_2 \mathbf{r}_1 (\mathcal{L}_i - \mathbf{k}_i) + \theta_3 \mathbf{r}_2 (\mathcal{G} - \mathbf{k}_i) \tag{3.13}$$

$$\mathbf{k}_i = \mathbf{k}_i + \boldsymbol{\Delta}_i \tag{3.14}$$

where $\boldsymbol{\Delta}_i$ denotes the perturbation to the vector of parameters $\mathbf{k}_i$ for particle i. $(\theta_1, \theta_2, \theta_3)$ are adjustable parameters, $\mathcal{L}_i$ denotes the best local solution found by particle i, and $\mathcal{G}$ denotes the best solution found over the entire population of particles. The quantities r_1 and r_2 denote uniform random vectors with the same dimension as the number of unknown model parameters ($\mathcal{K} \times 1$). In this study, we used $(\theta_1, \theta_2, \theta_3) = (1.0, 0.05564, 0.02886)$. The quality of parameter estimates was measured using two criteria, goodness of fit (model residual) and angle between the estimated parameter vector $\mathbf{k}_j$ and the true parameter set $\mathbf{k}^*$:

$$\alpha_j = cos^{-1}\left(\frac{\mathbf{k}_j \cdot \mathbf{k}^*}{\mathbf{k}_j \mathbf{k}^*}\right) \tag{3.15}$$

If the candidate parameter set $\mathbf{k}_j$ were perfect, the residual between the model and synthetic data and the angle between $\mathbf{k}_j$ and the true parameter set $\mathbf{k}^*$ would be equal to zero.

We modified our PSO implementation to simultaneously search over kinetic parameters and putative model control structures. In the combined case, each particle potentially carried a different model realization in addition to a different kinetic parameter vector. We kept the update rules the same (along with the update parameters). Thus, each particle competed on the basis of goodness of fit, which allowed different model structures to contribute to the overall behavior of the swarm. We considered five possible model structures (A through E), where network A was the correct formulation (used to generate

the synthetic data). We considered a population of 100 particles, where each particle in the swarm was assigned a model structure, and a random parameter vector. The PSO algorithm, model equations, and the objective function were encoded and solved in the Octave programming language (v 3.8.1; www.octave.org).

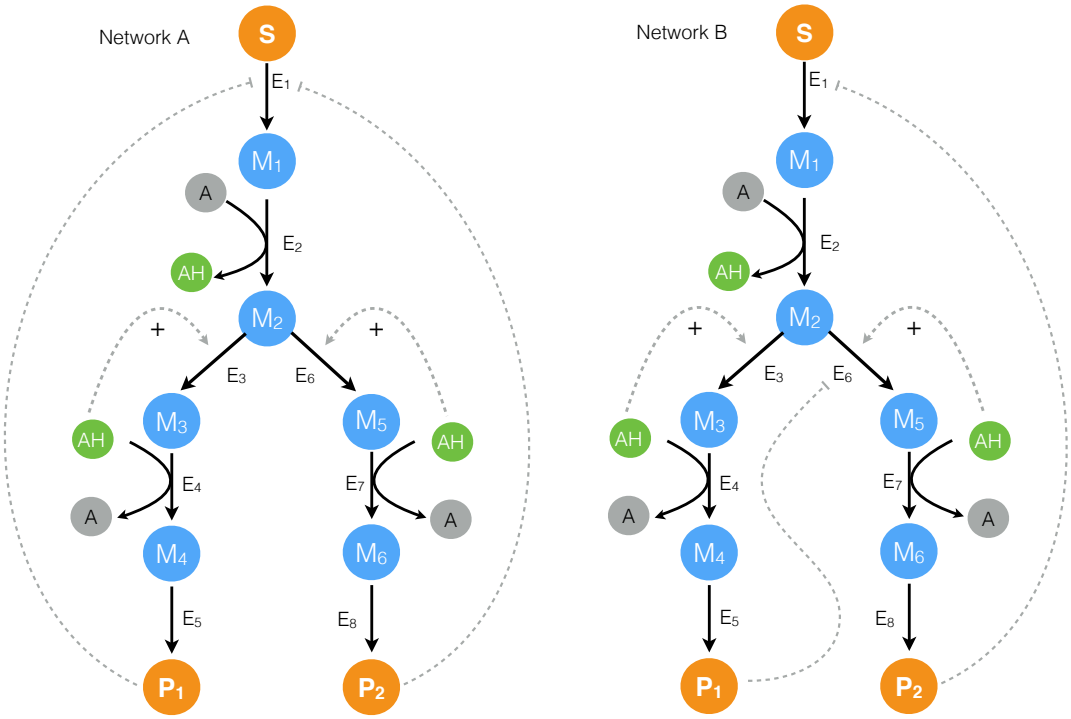

Figure 3.1: Proof-of-concept cell-free metabolic networks considered in this study. Substrate S is converted to products P_1 and P_2 through a series of chemical conversions catalyzed by enzyme(s) E_j. The activity of the pathway enzymes is subject to both positive and negative allosteric regulation.

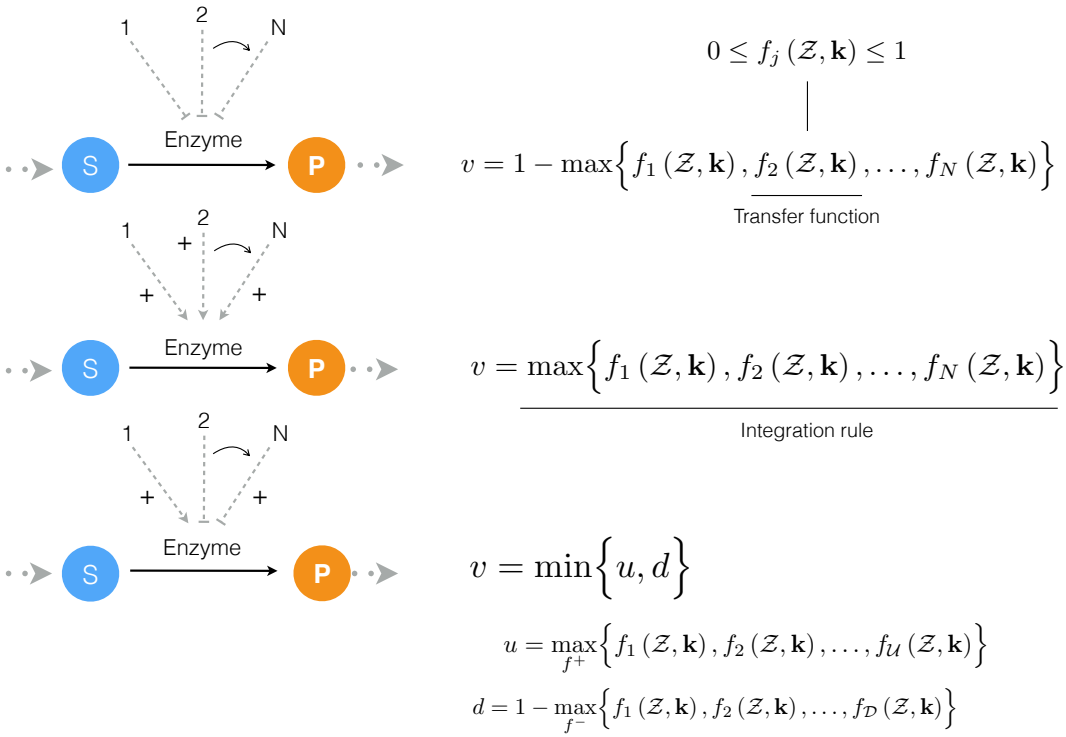

Figure 3.2: Schematic of rule-based allosteric enzyme activity control laws. Traditional enzyme kinetic expressions, e.g., Michaelis–Menten or multiple saturation kinetics, are multiplied by an enzyme activity control variable $0 \leq v_j \leq 1$. Control variables are functions of many possible regulatory factors encoded by arbitrary functions of the form $0 \leq f_j(\mathcal{Z}) \leq 1$. At each simulation time step, the v_j variables are calculated by evaluating integration rules such as the max or min of the set of factors $f_1, \ldots$ influencing the activity of enzyme E_j.

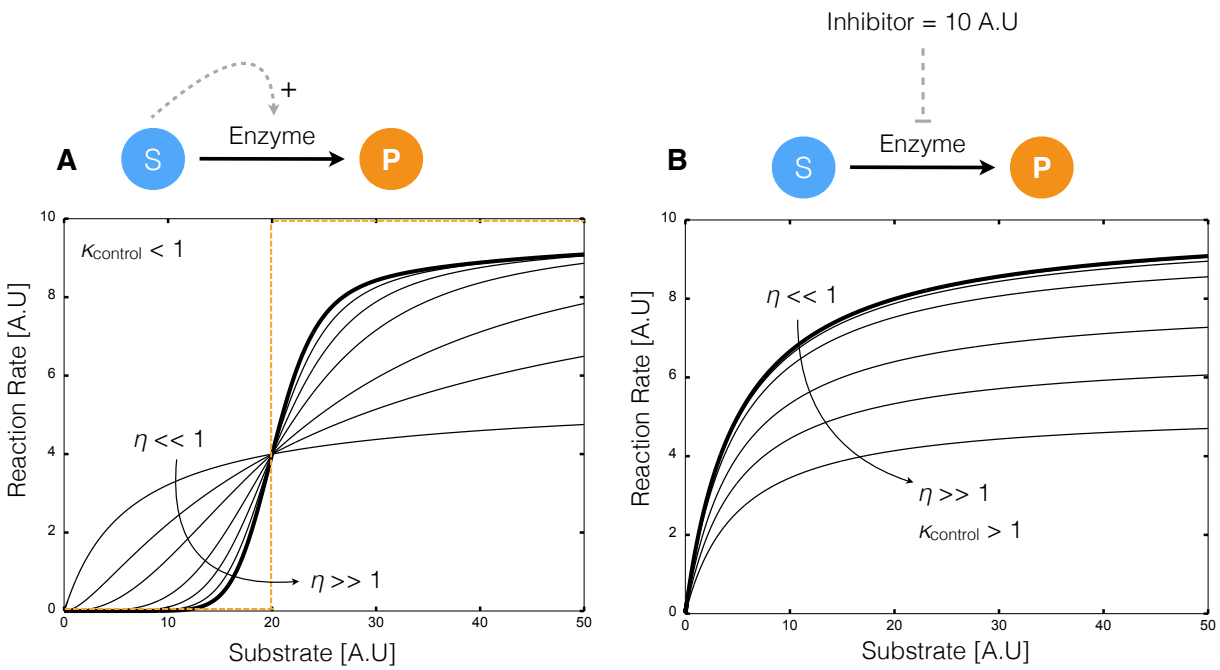

Figure 3.3: Kinetics of simple transformations in the presence of activation and inhibition. **A**:The conversion of substrate S to product P by enzyme E was activated by S. For a fixed control gain parameter $\kappa_{control}$, the reaction rate approached a step for increasing cooperativity control parameter η. For activation simulations $\kappa_{control}$ =0.05 and $\eta = \{0.01, 0.1, 1, 2, 4, 6, 8, 10\}$. **B**:The conversion of substrate S to product P by enzyme E with inhibitor I. For a fixed control gain parameter $\kappa_{control}$, the reaction rate approximated non-competitive inhibition for increasing cooperativity control parameter η. For the inhibition simulations $\kappa_{control}$ =1.5 and $\eta = \{0.01, 0.1, 1, 2, 4, 6, 8, 10\}$.

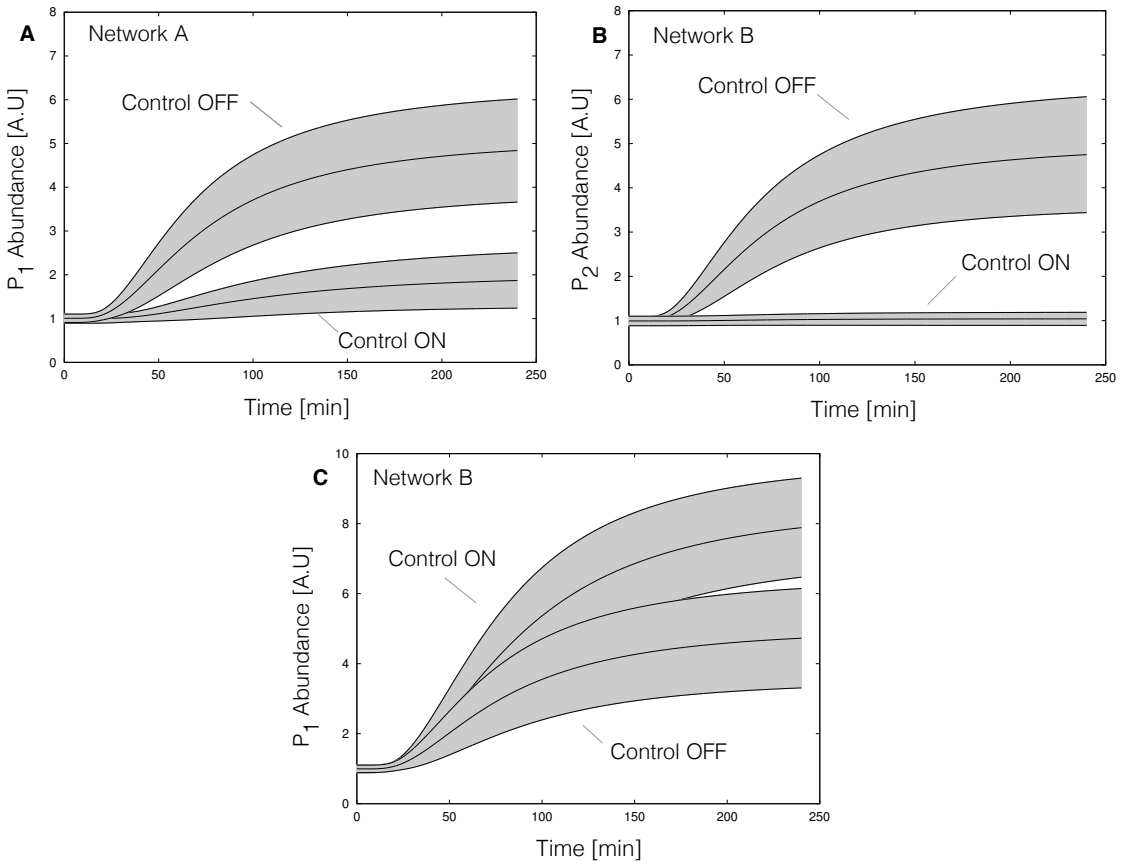

Figure 3.4: ON/OFF control simulations for Network A and Network B for an ensemble of 100 kinetic parameter sets versus time. For each case, simulations were conducted using kinetic and initial conditions generated randomly from a hypothetical true parameter set. The gray area represents ± one standard deviation surrounding the mean. Control parameters were fixed during the ensemble calculations. **A**: End product P_1 abundance versus time for Network A. The abundance of P_1 decreased with end product inhibition of E_1 activity (Control-ON) versus the no inhibition case (Control-OFF). **B**: End product P_2 abundance versus time for Network B. Inhibition of branch point E_6 by end product P_1 decreased P_2 abundance (Control-ON) versus the no inhibition case (Control-OFF). **C**: End product P_1 abundance versus time for Network A. Inhibition of branch point E_6 by end product P_1 decreased P_1 abundance (Control-ON) versus the no inhibition case (Control-OFF).

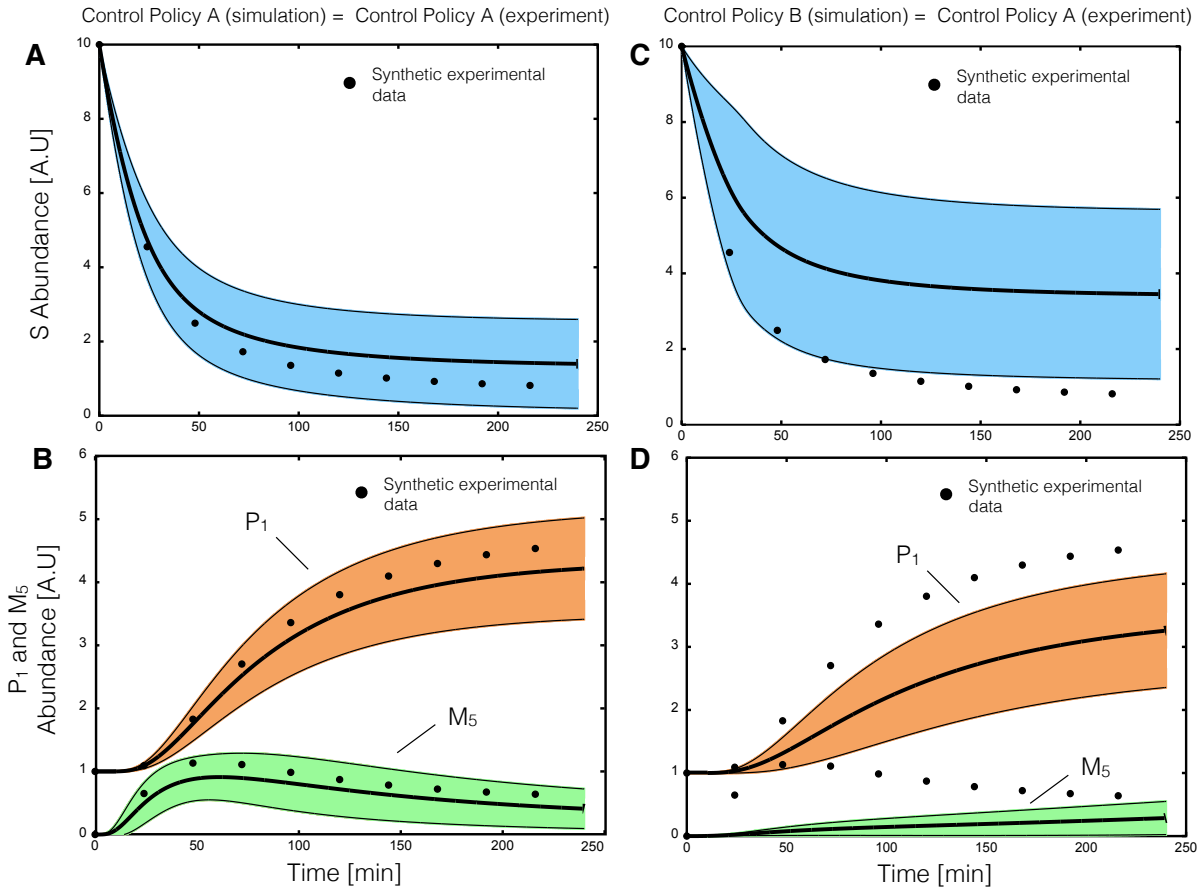

Figure 3.5: Parameter estimation from synthetic data for the same and mismatched allosteric control logic using particle swarm optimization (PSO). Synthetic experimental data was generated from a hypothetical parameter set using Network A, where substrate S, end product P_1 and intermediate M_5 were sampled approximately every 20 minutes. For cases **A,B** 20 particles were initialized with randomized parameters and allowed to search for 300 iterations. **A,B**: PSO estimated an ensemble of 20 parameters sets consistent with the synthetic experimental data assuming the correct enzymatic and control connectivity starting from randomized initial parameters. **C,D**: In the presence of control mismatch (Network B control policy simulated with Network A kinetic parameters) the ensemble of models did not describe the synthetic data. The synthetic data plotted here was unperturbed by noise. However, we assumed a constant coefficient of variation of 10% for the synthetic data during parameter estimation.

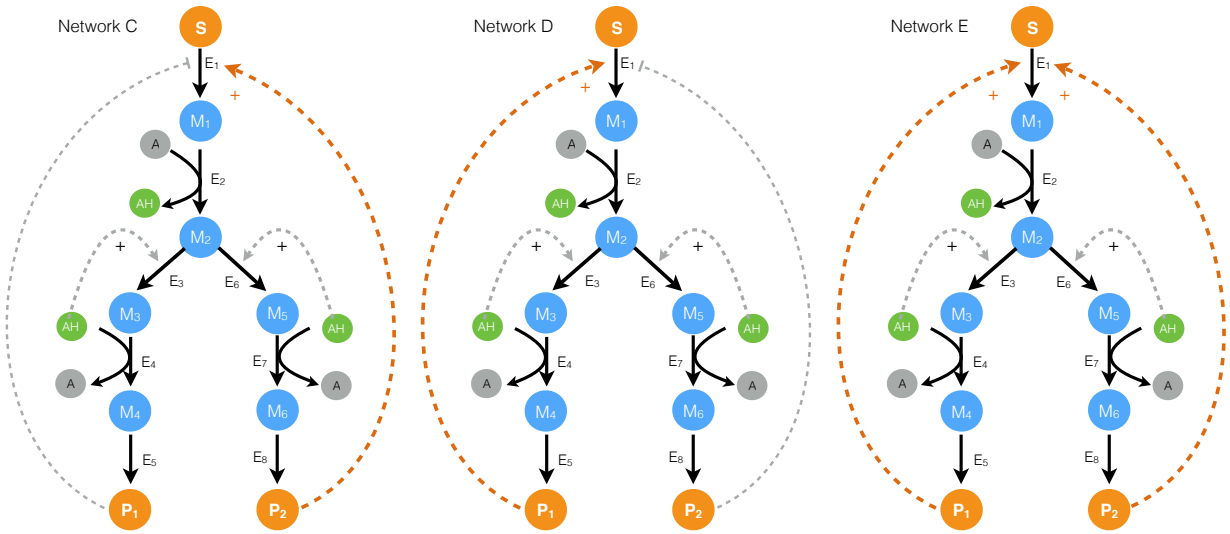

Figure 3.6: Schematic of the alternative allosteric control programs used in the structural particle swarm computation. Each network had the same enzymatic connectivity, initial conditions and kinetic parameters, but alternative feedback control structures for the first enzyme in the pathway.

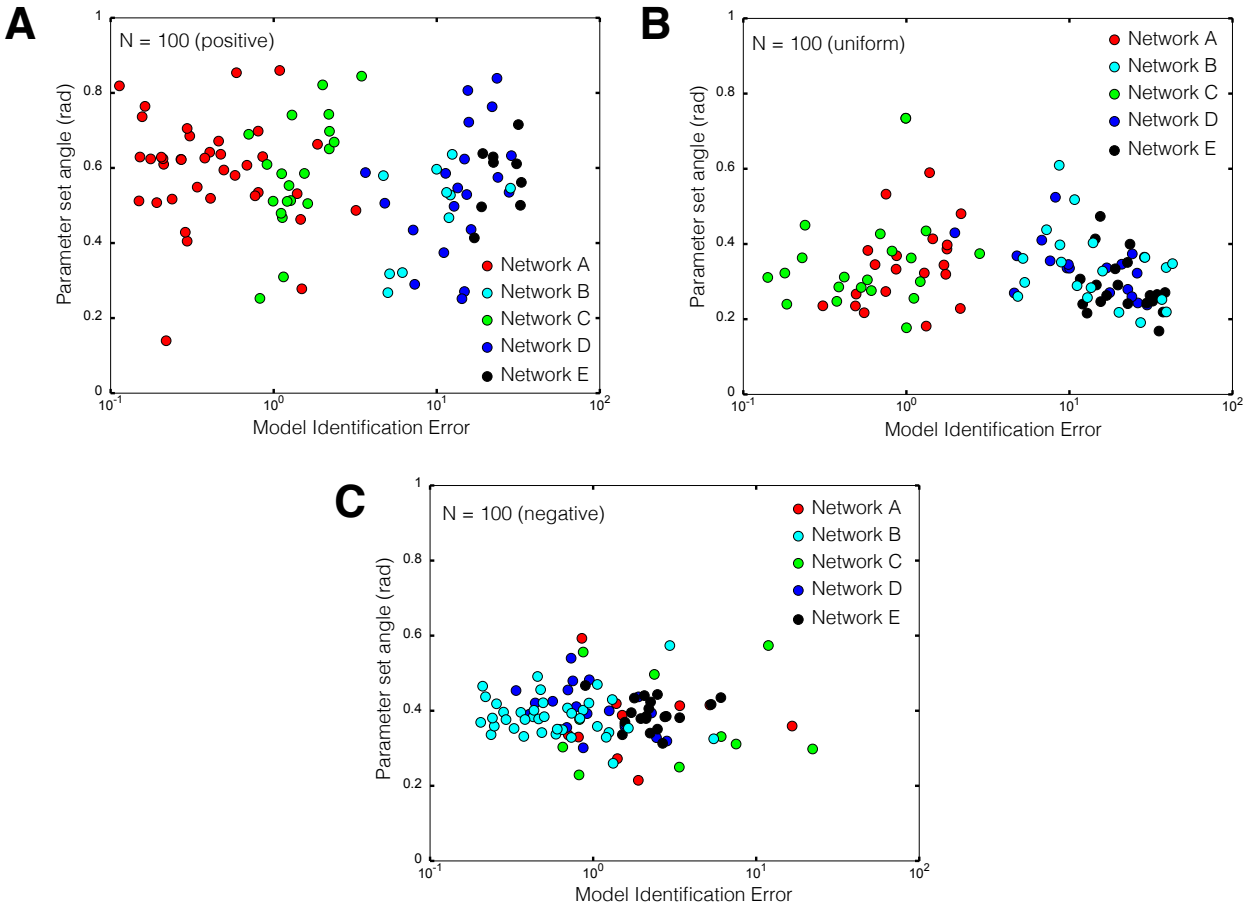

Figure 3.7: Combined control and kinetic parameter search using modified particle swarm optimization (PSO). A population of 100 particles was initialized with randomized kinetic parameters and one of five possible control configurations (Network A - E). Simulation error was minimized for a synthetic data set (S, end product P_1 and intermediate M_5 sampled approximately every 20 min) generated using Network A. **A**: Simulation error versus parameter set angle for 100 particles biased toward the correct regulatory program (A,B,C,D,E) = (40%, 10%, 20%, 20% and 10%). **B**: Simulation error versus parameter set angle for 100 uniformly distributed particles (A,B,C,D,E) = (20%, 20%, 20%, 20% and 20%). **C**: Simulation error versus parameter set angle for 100 negatively biased particles (A,B,C,D,E) = (10%, 40%, 10%, 20% and 20%). Network A (the correct structure) was preferentially identified for positively and uniform biased particle distributions, but misidentified in the presence of a large incorrect bias.

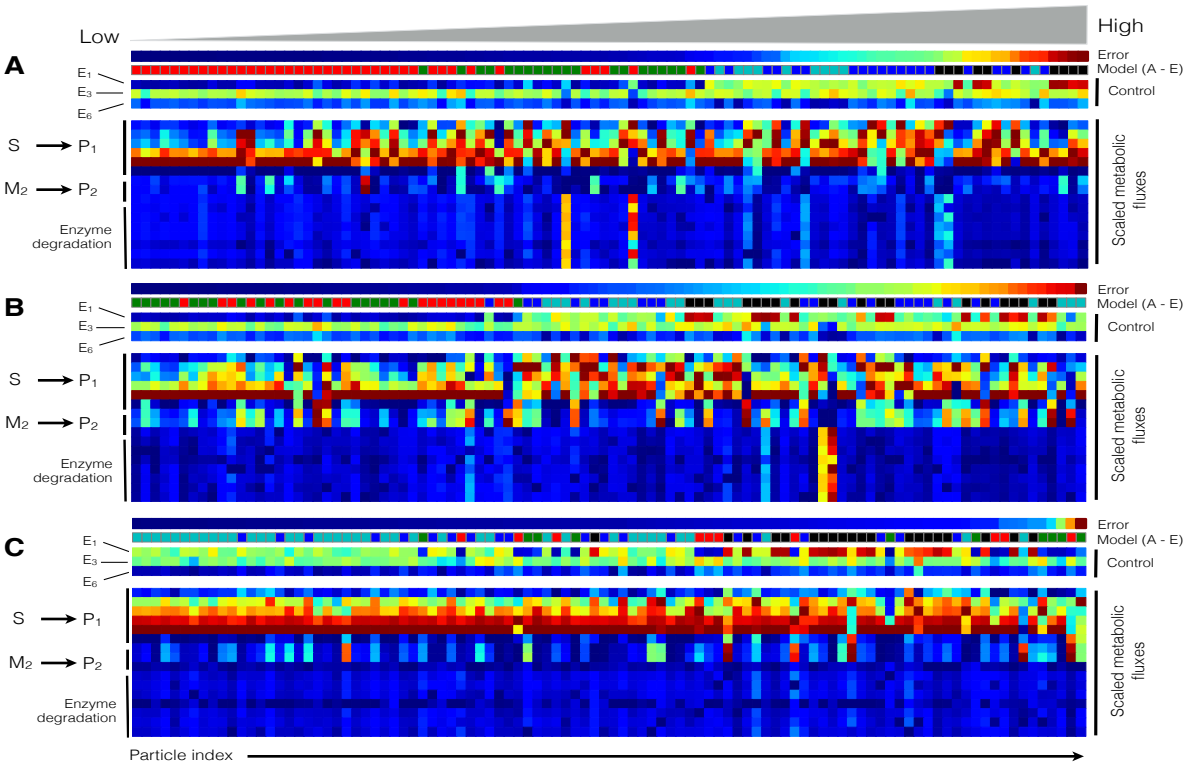

Figure 3.8: Metabolic flux and control variables as a function of network type and particle index at t = 100 min. The particle error, the control variables governing E_1, E_3 and E_6 activity (v_1, v_3 and v_3) and the scaled metabolic flux were calculated for the positively (top), uniformly (middle) and negatively (bottom) biased particle swarms (N = 100). Blue denotes a low value, while red denotes a high value for the respective quantity being plotted. The particles from each swarm were sorted based upon simulation error (low to high error). **A**: Model performance for the positively biased particle swarm as a function of particle index. **B**: Model performance for the uniformly biased particle swarm as a function of particle index. **C**: Model performance for the negatively biased particle swarm as a function of particle index. Models with significant control mismatch showed distinct control and flux patterns versus those models with the correct or closely related control policies. In particular, models with the correct control policy showed stronger inhibition of E_1 activity, leading to decreased flux from S→P_1. Conversely, models with significant mismatch had increased E_1 activity, leading to an altered flux distribution. This is especially apparent in the negatively biased particle swarm.

CHAPTER 4

REDUCED ORDER MODELING AND ANALYSIS OF THE HUMAN COMPLEMENT SYSTEM

Abstract

Complement [1] is an important pathway in innate immunity, inflammation, and many disease processes. However, despite its importance, there are few validated mathematical models of complement activation. In this study, we developed an ensemble of experimentally validated reduced order complement models. We combined ordinary differential equations with logical rules to produce a compact yet predictive model of complement activation. The model, which described the lectin and alternative pathways, was an order of magnitude smaller than comparable models in the literature. We estimated an ensemble of model parameters from *in vitro* dynamic measurements of the C3a and C5a complement proteins. Subsequently, we validated the model on unseen C3a and C5a measurements not used for model training. Despite its small size, the model was surprisingly predictive. Global sensitivity and robustness analysis suggested complement was robust to any single therapeutic intervention. Only the simultaneous knockdown of both C3 and C5 consistently reduced C3a and C5a formation from all pathways. Taken together, we developed a validated mathematical model of complement activation that was computationally inexpensive, and could easily be incorporated into pre-existing or new pharmacokinetic models of immune system function. The model described experimental data, and predicted the need for multiple points of therapeutic intervention to fully disrupt complement activation.

[1] Adapted with permission from Sagar A, Dai W, Minot M, Varner JD (2016) "Reduced order modeling and analysis of the human complement system" PLoS ONE *Submitted*

4.1 Introduction

Complement is an important pathway in innate immunity. It plays a significant role in inflammation, host defense as well as many disease processes. Complement was discovered in the late 1880s where it was found to 'complement' the bactericidal activity of natural antibodies [131]. However, research over the past decade has suggested the importance of complement extends beyond innate immunity. For example, complement contributes to tissue homeostasis [132]. It has also has been linked with several diseases including Alzheimers, Parkinson's, multiple sclerosis, schizophrenia, rheumatoid arthritis and sepsis [10, 133]. Complement also plays positive and negative roles in cancer; attacking tumor cells with altered surface proteins in some cases, while potentially contributing to tumor growth in others [134, 135]. Lastly, several other important biochemical systems are integrated with complement including the coagulation cascade, the autonomous nervous system and inflammation [135]. Thus, complement is important in a variety of beneficial and potentially harmful functions in the body. Despite its importance, there have been few approved complement specific therapeutics, largely because of safety concerns and challenging pharmacokinetic constraints, however, progress is being made [136].

The complement cascade involves many soluble and cell surface proteins, receptors and regulators [137, 138]. The outputs of complement are the Membrane Attack Complex (MAC), and the inflammatory mediator proteins C3a and C5a. The membrane attack complex, generated during the terminal phase of the response, forms transmembrane channels which disrupt the membrane integrity of targeted cells, leading to cell lysis and death. On the other hand, the C3a and C5a proteins act as a bridge between innate and adaptive immunity, and play an important role in regulating inflammation [134]. Complement activation takes places through three pathways: the classical, the lectin and the

alternate pathways. The classical pathway is triggered by antibody recognition of foreign antigens or other pathogens. A multimeric protein complex C1 binds antibody-antigen complexes and undergoes a conformational change, leading to an activated form with proteolytic activity. The activated C1-complex cleaves soluble complement proteins C4 and C2 into C4a, C4b, C2a and C2b, respectively. The C4a and C2b fragments bind to form the C4bC2a protease, also known as the classical pathway C3 convertase (CP C3 convertase). The lectin pathway is initiated through the binding of L-ficolin or Mannose Binding Lectin (MBL) to carbohydrates on the surfaces of bacterial pathogens. These complexes, in combination with mannose-associated serine proteases 1 and 2 (MASP-1/2), also cleave C4 and C2, leading to additional CP C3 convertase. Thus, the classical and lectin pathways, initiated by different cues on foreign surfaces, converge at the CP C3 convertase. On the other hand, the alternate pathway is activated by a 'tickover' mechanism in which complement protein C3 is spontaneously hydrolyzed to form an activated intermediate C3w; C3w recruits factor B and factor D, leading to the formation of C3wBb. C3wBb cleaves C3 into C3a and C3b, where the C3b fragment further recruits additional factor B and factor D to form C3bBb, the alternate C3 convertase (AP C3 convertase) [139]. The role of classical and alternate C3 convertases is varied. First, AP C3 convertases mediate signal amplification. AP C3 convertases cleave C3 into C3a and C3b; the C3b fragment is then free to form additional alternate C3 convertases, thereby forming a positive feedback loop. Next, AP/CP C3 convertases link complement initiation with the terminal phase of the cascade through the formation of C5 convertases. Both classical and alternate C3 convertases can recruit C3b subunits to form the classical pathway C5 convertase (C4bC2aC3b, CP C5 convertase), and the alternate pathway C5 convertase (C3bBbC3b, AP C5 convertase), respectively. Both C5 convertases cleave C5 into the C5a and C5b fragments. The C5b fragment, along with the complement proteins C6, C7, C8 and multiple C9s, form the membrane attack complex. On the other hand, both C3a and

C5a are important inflammatory signals involved in several responses [137, 138]. Thus, the complement cascade attacks invading pathogens, while acting as a beacon for adaptive immunity.

The complement cascade is regulated by plasma and host cell surface proteins which balance host safety with effectiveness. The initiation of the classical pathway via complement protein C1 is controlled by the C1 Inhibitor (C1-Inh); C1-Inh irreversibly binds to and deactivates the active subunits of C1, preventing chronic complement activation [140]. Regulation of upstream processes in the lectin and alternate pathways also occurs through the interaction of the C4 binding protein (C4BP) with C4b, and factor H with C3b [141]. Interestingly, both factor H and C4BP are capable of binding their respective targets while in convertase complexes as well. At the host cell surface, membrane cofactor protein (MCP or CD46) can interact with C4b and C3b, which protects the host cell from complement self-activation [142]. Delay accelerating factor (DAF or CD55) also recognizes and dissociates both C3 and C5 convertases on host cell surfaces [143]. More generally the well known inflammation regulator Carboxypeptidase-N has broad activity against the complement proteins C3a, C4a, and C5a, rendering them inactive by cleavage of carboxyl-terminal arginine and lysine residues [144]. Although Carboxypeptidase-N does not directly influence complement activation, it silences the important inflammatory signals produced by complement. Lastly, assembly of the MAC complex itself can be inhibited by vitronectin and clusterin in the plasma, and CD59 at the host surface [145, 146]. Thus, there are many points of control which influence complement across the three activation pathways.

Developing quantitative mathematical models of complement could be crucial to fully understanding its role in the body. Traditionally, complement models have been formulated as systems of linear or non-linear ordinary differential equations (ODEs). For ex-

ample, Hirayama et al., modeled the classical complement pathway as a system of linear ODEs [147], while Korotaevskiy and co-workers modeled the classical, lectin and alternate pathways as a system of non-linear ODEs [148]. More recently, large mechanistic models of sections of complement have also been proposed. For example, Liu et al., analyzed the formation of the classical and lectin C3 convertases, and the regulatory role of C4BP using a system of 45 non-linear ODEs with 85 parameters [149]. Zewde and co-workers constructed a detailed mechanistic model of the alternative pathway which consisted of 107 ODEs and 74 kinetic parameters and delineated between the fluid, host and pathogen surfaces [146]. However, these previous studies involved large models with little experimental validation. Thus, while these models are undoubtably important theoretical tools, it is unclear if they can describe or quantitatively predict complement measurements. The central challenge of complement model identification is the estimation of model parameters from experimental measurements. Unlike other important cascades, such as coagulation where there are well developed experimental tools and publicly available data sets, the data for complement is relatively sparse. Data sets with missing or incomplete data, and limited dynamic data also make the identification of large mechanistic complement models difficult. Thus, reduced order approaches which describe the biology of complement using a limited number of species and parameters could be important for pharmacokinetic model development, and for our understanding of the varied role of complement in the body.

4.2 Results

In this study, we estimated an ensemble of experimentally validated reduced order complement models using multiobjective optimization. The modeling approach combined ordinary differential equations with logical rules to produce a complement model with a limited number of equations and parameters. The reduced order model, which described the lectin and alternative pathways, consisted of 18 differential equations with 28 parameters. Thus, the model was an order of magnitude smaller and included more pathways than comparable models in the literature. We estimated an ensemble of model parameters from *in vitro* time series measurements of the C3a and C5a complement proteins. Subsequently, we validated the model on unseen C3a and C5a measurements not used for model training. Despite its size, the model was surprisingly predictive. After validation, we performed global sensitivity and robustness analysis to estimate which parameters and species controlled model performance. Sensitivity analysis suggested CP C3 and C5 convertase parameters were critical, while robustness analyses suggested complement was robust to any single therapeutic intervention; only the knockdown of both C3 and C5 consistently reduced C3a and C5a formation for all cases. Taken together, we developed a reduced order complement model that was computationally inexpensive, and could easily be incorporated into pre-existing or new pharmacokinetic models of immune system function. The model described experimental data, and predicted the need for multiple points of intervention to disrupt complement activation.

4.2.1 Reduced order complement network.

The complement model described the alternate and lectin pathways (Fig. 4.1). A trigger event initiated the lectin pathway (encoded as a logical rule), which activated the cleavage of C2 and C4 into C2a, C2b, C4a and C4b, respectively. Classical Pathway (CP) C3 convertase (C4aC2b) then catalyzed the cleavage of C3 into C3a and C3b. The alternate pathway was initiated through the spontaneous hydrolysis of C3 into C3a and C3b. The C3b fragments generated by hydrolysis (or by CP C3 convertase) could then form the alternate pathway (AP) C3 convertase (C3bBb). We did not consider C3w, nor the formation of the initial alternate C3 convertase (C3wBb). Rather, we assumed C3w was equivalent to C3b and only modeled the formation of the main AP C3 convertase. Both the CP and AP C3 convertases catalyzed the cleavage of C3 into C3a and C3b. A second C3b fragment could then bind with either the CP or AP C3 convertase to form the CP or AP C5 convertase (C4bC2aC3b or C3bBbC3b). Both C5 convertases catalyzed the cleavage of C5 into the C5a and C5b fragments. In this study, we simplified the model by assuming both factor B and factor D were in excess. However, we did explicitly account for the action of two other control proteins, factor H and C4BP. Lastly, we did not consider MAC formation, instead we stopped at C5a and C5b. Lectin pathway activation, and C3/C5 convertase activity were modeled using a combination of saturation kinetics and non-linear transfer functions, which resulted in a significant size reduction of the model, while maintaining performance. Binding interactions were modeled using mass-action kinetics, where we assumed all binding was irreversible. Thus, while the reduced order complement model encoded significant biology, it was highly compact consisting of only 18 differential equations and 28 model parameters. Next, we estimated an ensemble of model parameters from time series measurements of the C3a and C5a complement proteins.

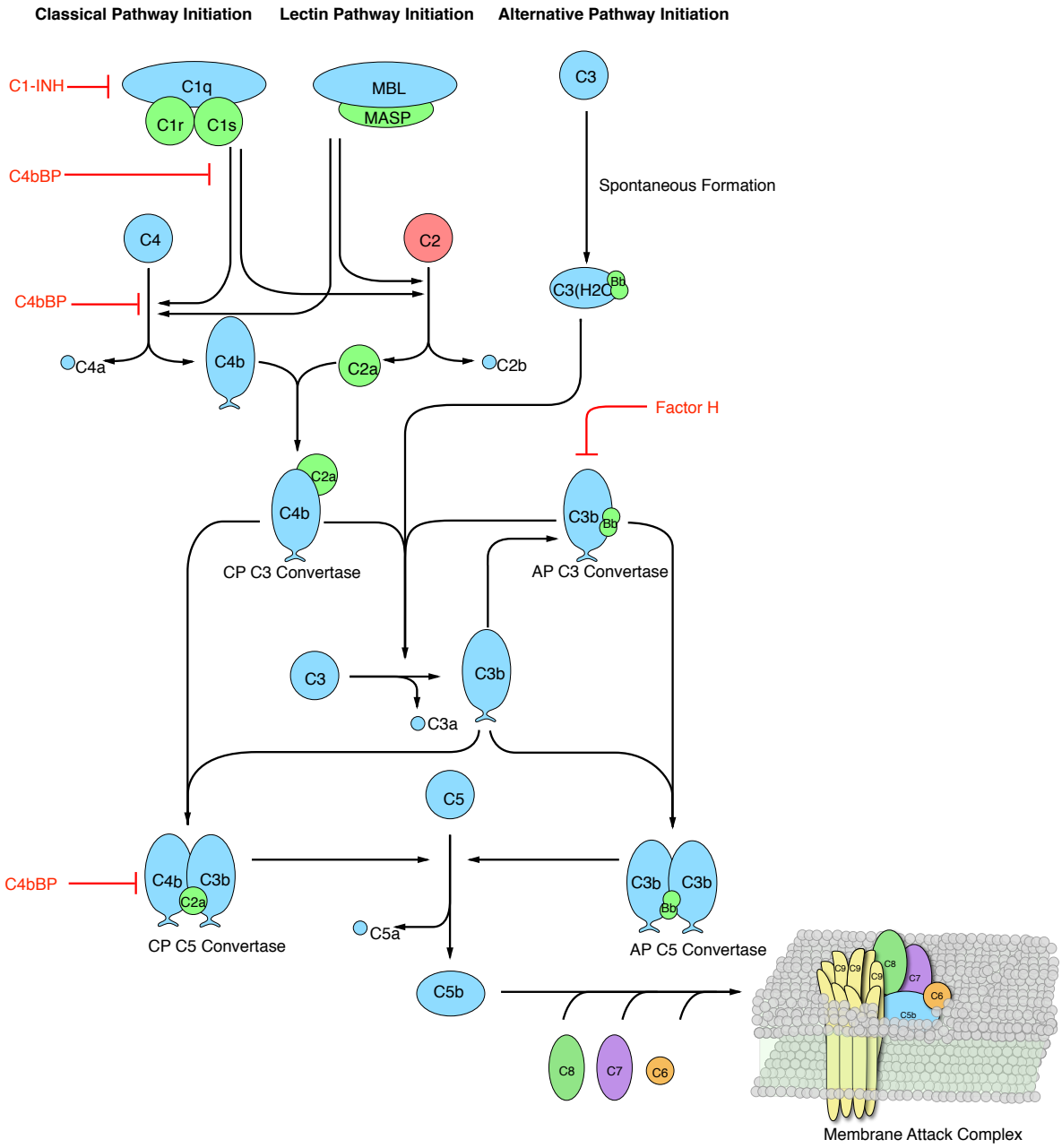

Figure 4.1: Simplified schematic of the human complement system. The complement cascade is activated through three pathways: the classical, the lectin, and the alternate pathways. Complement initiation results in the formation of classical or alternative C3 convertases, which amplify the initial complement response and signal to the adaptive immune system by cleaving C3 into C3a and C3b. C3 convertases further react to form C5 convertases which catalyze the cleavage of the C5 complement protein to C5a and C5b. C5b is critical to the formation of the membrane attack complex (MAC), while C5a recruits an adaptive immune response.

4.2.2 Estimating an ensemble of reduced order complement models.

A critical challenge for the development of any dynamic model is the estimation of model parameters. We estimated an ensemble of complement model parameters using *in vitro* time-series data sets generated with and without zymosan, a lectin pathway activator [4]. The residual between model simulations and experimental measurements was minimized using the Pareto Optimal Ensemble Technique (JuPOETs) [150] starting from a initial guess generated by the dynamic optimization with particle swarms (DOPS) routine. Unless otherwise specified, all initial conditions were assumed to be at their mean physiological values. While we had significant training data, the parameter estimation problem was underdetermined (we were not able to uniquely determine model parameters). Thus, instead of using the best-fit yet uncertain parameter set, we estimated an ensemble of probable parameter sets to quantify model uncertainty (N = 2100, see materials and methods). The complement model ensemble captured the behavior of both the alternate and lectin pathways (Fig. 4.2). To estimate alternate pathway model parameters, we used C3a and C5a measurements in the absence of zymosan (Fig. 4.2A and B). On the other hand, lectin pathway parameters were estimated from C3a and C5a measurements in the presence of 1mg/ml zymosan (Fig. 4.2C and D). The reduced order model reproduced a panel of alternate and lectin pathway data sets in the neighborhood of physiological factor and inhibitor concentrations. The model fit for parameter sets estimated by JuPOETs, quantified by the Akaike information criterion (AIC), was statistically significantly different than a random parameter control for each case at a 95% confidence level. However, it was unclear whether the reduced order model could predict new data, without updating the model parameters. To address this question, we fixed the model parameters and simulated data sets not used for model training.

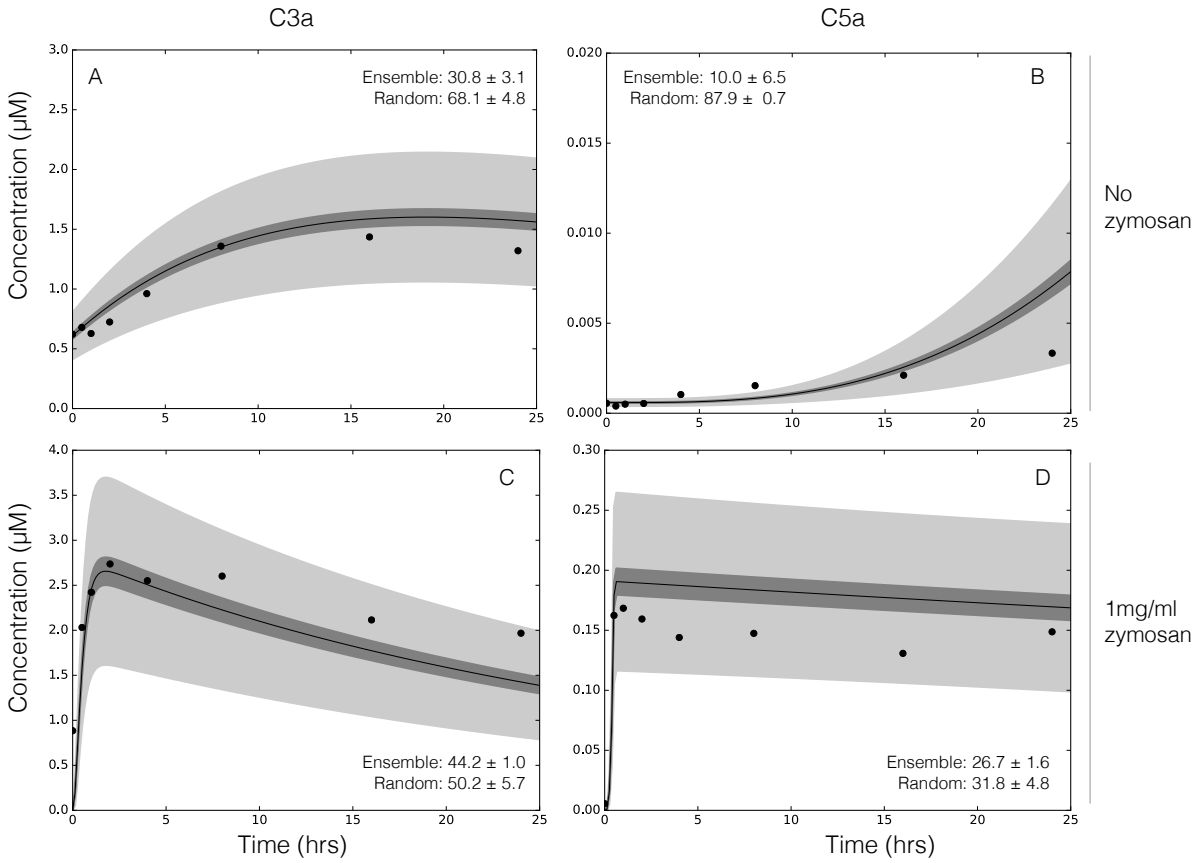

Figure 4.2: Reduced order complement model training. An ensemble of model parameters were estimated using multiobjective optimization from C3a and C5a measurements with and without zymosan [4]. The model was trained using C3a and C5a data generated from the alternative pathway (**A**–**B**) and lectin pathway initiated with 1 mg/ml zymosan (**C**–**D**). The solid black lines show the simulated mean value of C3a or C5a for the ensemble, while the dark shaded region denotes the 99% confidence interval of mean. The light shaded region denotes the 99% confidence interval of the simulated C3a and C5a concentration. All initial conditions were assumed to be at their physiological serum levels unless otherwise noted.

We tested the predictive power of the reduced order complement model with data not used during model training (Fig. 4.3). Six validation cases were considered, three for C3a and C5a each, respectively. Similar to model training, we compared the AIC for each prediction case to a randomized parameter family. All model parameters and initial conditions were fixed for the validation simulations (with the exception of zymosan, and other experimentally mandated changes). The ensemble of reduced order models predicted the qualitative dynamics of C3a formation (Fig. 4.3, top), and C5a formation (Fig. 4.3, bottom) at three inducer concentrations. For each training case, the AIC was statistically significantly different than the random parameter control for a 95% confidence level. The rate of C3a formation and C3a peak time were directly proportional to initiator dose. Similarly, the C5a plateau and rate of formation were also directly proportional to initiator dose, with the lag time being indirectly proportional to initiator exposure for both C3a and C5a. However, there were shortcomings with model performance. First, while the overall C3a trend was captured (within the 99% confidence interval), the C3a dynamics were too fast with the exception of the low dose case. We believe the C3a time scale was related to our choice of training data, how we modeled the tickover mechanism, and factor B and D limitation. We trained the model using either no or 1 mg/ml zymosan, but predicted cases in a different initiator range; comparing training to prediction, the model performance e.g., the shape of the C3a trajectory was biased towards either high or very low initiator doses. Next, tickover was modeled as a first-order generation processes where C3wBb formation and activity was lumped into the AP C3 convertase. Thus, we skipped an important upstream step which could influence AP C3 convertase formation by attenuating the rate C3 cleavage into C3a and C3b. We also assumed both factor B and factor D were not limiting, thereby artificially accelerating the rate of AP C3 convertase formation. The C5a predictions followed a similar trend as C3a; we captured the long-time C5a behavior but over predicted the time scale of C5 cleavage. However, because

the C5a time scale depends strongly upon C3 convertase formation, we can likely correct the C5 issues by fixing the rate of C3 cleavage. Despite these shortcomings, we qualitatively predicted experimental measurements not used for model training typically within the 99% confidence of the ensemble, for three inducer levels. Next, we used global sensitivity and robustness analysis to determine which parameters and species controlled the performance of the complement model.

4.2.3 Global analysis of the reduced order complement model.

We conducted sensitivity analysis to estimate which parameters controlled the performance of the reduced order complement model. We calculated the total sensitivity of the C3a and C5a residual to changes in model parameters with and without zymosan (Fig. 4.4). In the absence of zymosan (where only the alternative pathway is active), the most sensitive parameter was the rate constant governing the assembly of the AP C3 convertase, as well as the rate constant controlling basal C3b formation via the tickover mechanism. The C5a trajectory was sensitive to the AP C5 convertase kinetic parameters (Fig. 4.4A). Interestingly, neither the rate nor the saturation constant governing AP C3 convertase activity were sensitive in the absence of zymosan. Thus, C3a formation in the alternative pathway was more heavily influenced by the spontaneous hydrolysis of C3, rather than AP C3 convertase activity, in the absence of zymosan. In the presence of zymosan, the C3a residual was controlled by the formation and activity of the CP C3 convertase, as well as tickover and degradation parameters. On the other hand, the C5a residual was controlled by the formation and activity of CP C5 convertase, and tickover C3b formation in the presence of zymosan (Fig. 4.4B). The lectin initiation parameters were sensitive, but to a lesser extent than CP convertase kinetic parameters and

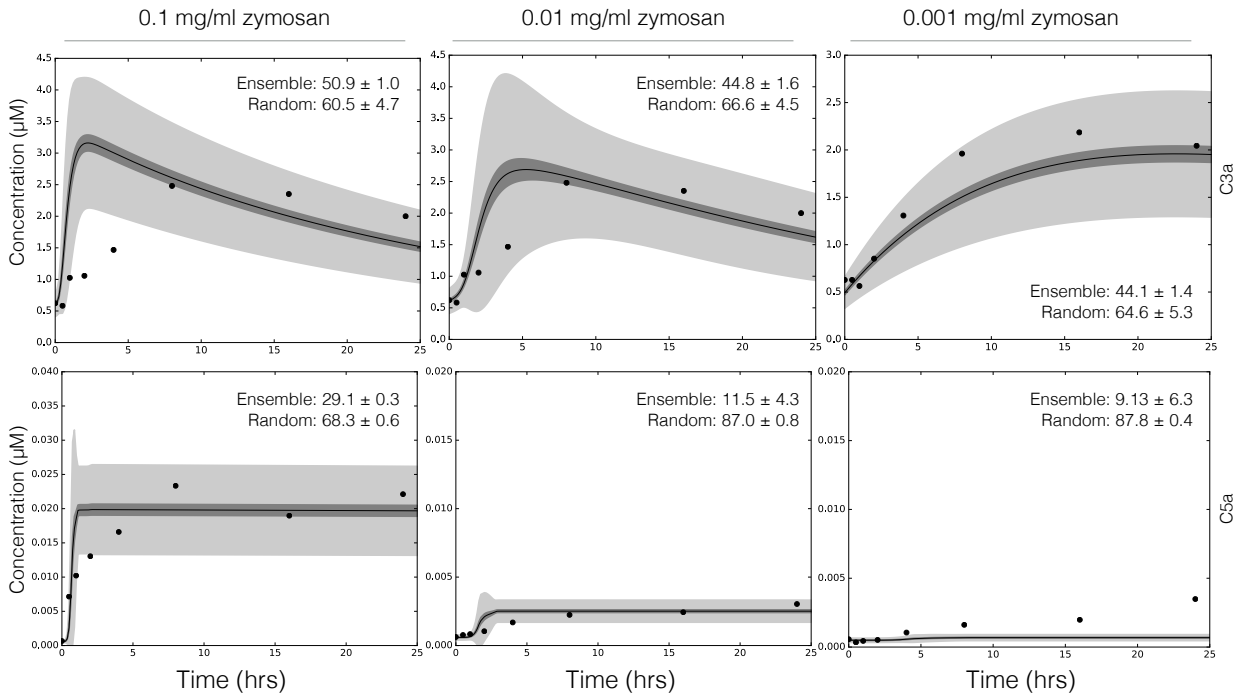

Figure 4.3: Reduced order complement model predictions. Simulations of C3a and C5a generated in the lectin pathway using 0.1 mg/ml, 0.01 mg/ml, and 0.001 mg/ml zymosan were compared with the corresponding experimental measurements. The solid black lines show the simulated mean value of C3a or C5a for the ensemble, while the dark shaded region denotes the 99% confidence interval of mean. The light shaded region denotes the 99% confidence interval of the simulated C3a and C5a concentration. All initial conditions were assumed to be at their physiological serum levels unless otherwise noted.

tickover C3b formation. Thus, sensitivity analysis suggested that CP C3/C5 convertase formation and activity dominated in the presence of zymosan, but tickover parameters and AP C5 convertase were more important without initiator. AP C3 convertase assembly was important, but its activity was not. Next, we compared the sensitivity results to current therapeutic approaches; pathways involving sensitive parameters have been targeted for clinical intervention (Fig. 4.4C). In particular, the sensitivity analysis suggested AP/CP C5 convertase inhibitors, or interventions aimed at attenuating C3 or C5 would most strongly influence complement performance. Thus, there was at least a qualitative overlap between sensitivity and the potential of biochemical efficacy. However, total sensitivity coefficients quantify how simultaneous changes in many parameters e.g., rate or saturation constants affect model performance (in this case model fit). To better understand the role of each parameter, and parameter combination, we explored how finite changes in parameter combinations influenced model performance.

Pairwise parameter perturbations identified crosstalk within the complement model (Fig. 4.5). We perturbed each pairwise combination of parameters by 10%, and calculated the distance between the perturbed and nominal state for each parameter set in the ensemble. We then clustered the mean response of each parameter combination based upon the euclidian distance between the perturbed and nominal states into low (green), medium (red) and high (blue) response clusters. A low response (white) meant the parameter perturbations did not significantly change the system state compared with the nominal case. Four of the 28 parameters (or approximately 14% of the overall model parameters) were in the high response cluster (Fig. 4.5, blue cluster). These parameters included the rate constant controlling the basal formation of C3b (#12), C3a degradation (#26) as well as the catalytic rate constant governing CP C3 convertase activity (#22). The only C5 related parameter in the high response group was the rate constant controlling

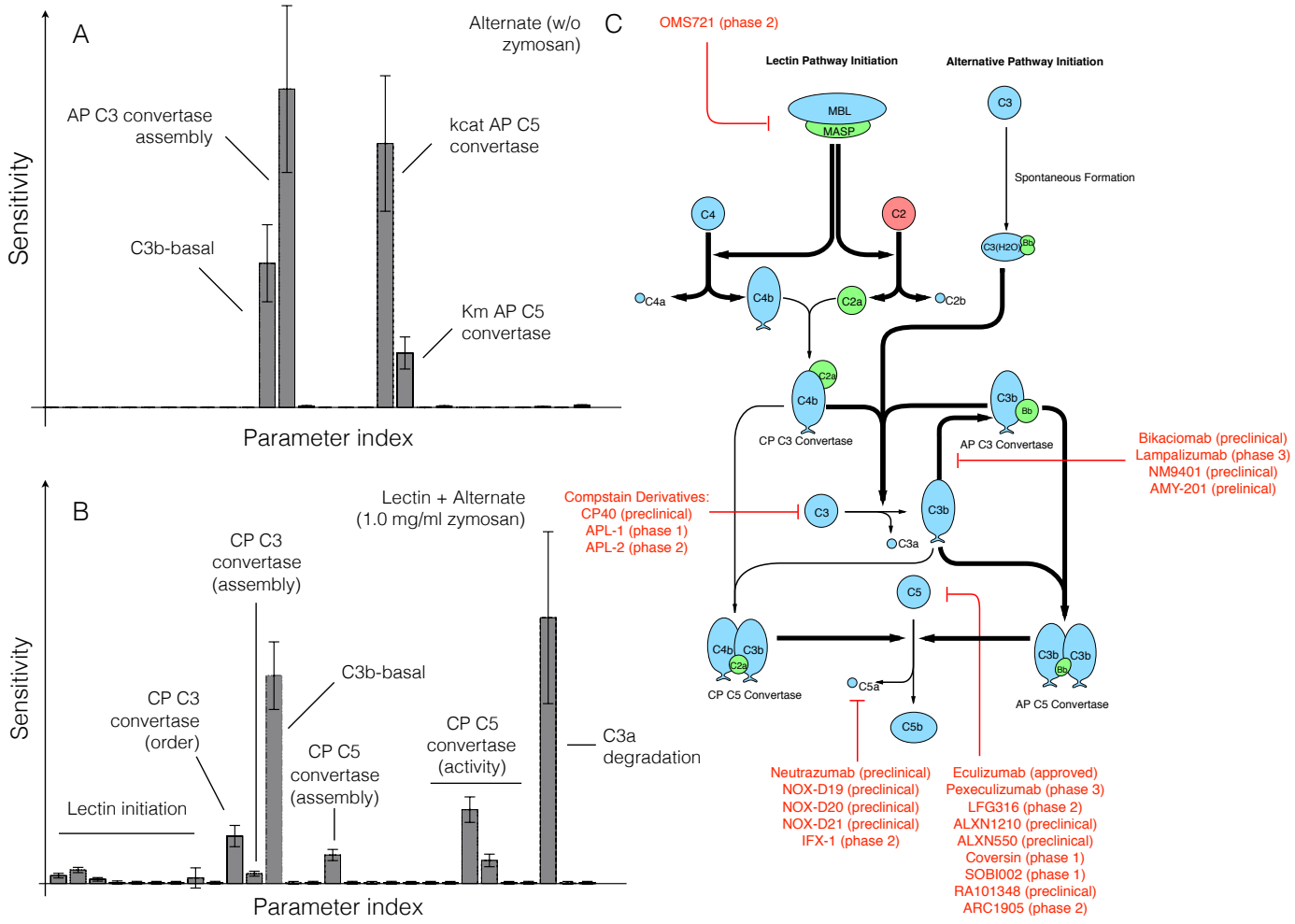

Figure 4.4: Global sensitivity analysis of the reduced order complement model. Sensitivity analysis was conducted on the two objectives used for model training. **A**: Sensitivity of the C3a and C5a residual w/o zymosan. **B**: Sensitivity of the C3a and C5a residual with 1 mg/ml zymosan. The bars denote the mean total sensitivity index for each parameter, while the error bars denote the 95% confidence interval. **C**: Pathways controlled by the sensitivity parameters. Bold black lines indicate the pathway involves one or more sensitive parameters, while the red lines show current therapeutics targets. Current complement therapeutics were taken from the review of Morgan and Harris [5].

the formation of CP C5 convertase (#15). Approximately, 36%, or 10 of the 28 model parameters, were clustered in the medium impact cluster (Fig. 4.5, red cluster). Three parameters (#10, #1, #27) were especially important in this cluster; The reaction order governing CP C3 convertase activity was important (#10), along with the rate constant controlling C4a and C4b formation from C4 in the lectin initiation pathway (#1), and the constant controlling the inhibitory action of C4BP (#27). Lastly, 50% of the model parameters were clustered in the low response cluster (Fig. 4.5, green cluster). Many of these parameters influenced complement activation; for example, parameter #23 (the CP C3 convertase saturation constant) was important, just not to the extent of other model parameters. Pairwise synergistic interactions between parameters were also identified. For example, in the high impact cluster, three synergistic relationships were identified, a single positive and two negative cases. Parameters #12 (rate constant governing basal C3b formation) and #15 (formation of CP C5 conevertase) acted synergistically to increase the system response. On the other hand, simultaneously changing parameters #12 and #22 or #15 and #26 decreased the system response relative to a single perturbation. However, the most striking examples of synergy occurred in the medium impact cluster; for example, simultaneously increasing parameters #13 (rate constant governing AP C3 convertase formation) and #19 (saturation constant governing AP C5 convertase activity) significantly changed the model state. Changes in parameter #3 (rate constant governing C2a and C2b formation from C2) showed both positive and negative synergistic effects depending upon the other parameter that was perturbed. Taken together, [FINISH ME]. However, sensitivity coefficients quantify how changes in parameters e.g., rate or saturation constants affect model performance. To more closely simulate a clinical intervention e.g., administration of anti-complement inhibitors, we performed robustness analysis in the absence and presence of flow.

Robustness analysis in the absence of flow suggested there was no single intervention that inhibited complement activation in the presence of both initiation pathways (Fig. 4.6). Robustness coefficients quantify the response of a protein to a macroscopic structural or operational perturbation to a biochemical network. Here, we computed how the C3a and C5a trajectories responded to a decrease in the initial abundance of C3 and/or C5 with and without lectin initiator. We simulated the addition of different doses of anti-complement inhibitor cocktails by decreasing the initial concentration of C3, C5 or the combination of C3 and C5 by 50%, 90% and 99%. This would be conceptually analogous to the administration of a C3 inhibitor e.g., Compstatin alone or combination with Eculizumab (Fig. 4.4C). The response of the complement model to different knock-down magnitudes was non-linear; a 90% knock-down had an order of magnitude more impact than a 50% knock-down. As expected, a C5 knockdown had no effect on C3a formation for either the alternate (Fig. 4.6A) or lectin pathways (Fig. 4.6B). However, C3a and to a greater extent C5a abundance decreased with decreasing C3 concentration in the alternate pathway (Fig. 4.6A). This agreed with the sensitivity results; changes in AP C3-convertase formation affected the downstream dynamics of C5a formation. Thus, if we only considered the alternate pathway, C3 alone could be a reasonable target, especially given that C5a formation was surprisingly robust to C5 levels in the alternate pathway. Yet, when both pathways were activated, C5a levels were robust to the initial C3 concentration (Fig. 4.6B); even 1% of the nominal C3 was able to generate enough AP/CP C5 convertase to maintain C5a formation. Thus, the only reliable intervention that consistently reduced both C3a and C5a formation for all cases was a knockdown of both C3 and C5. For example, a 90% decrease of both C3 and C5 reduced the formation of C5a by an order of magnitude, while C3a was reduced to a lesser extent (Fig. 4.6B).

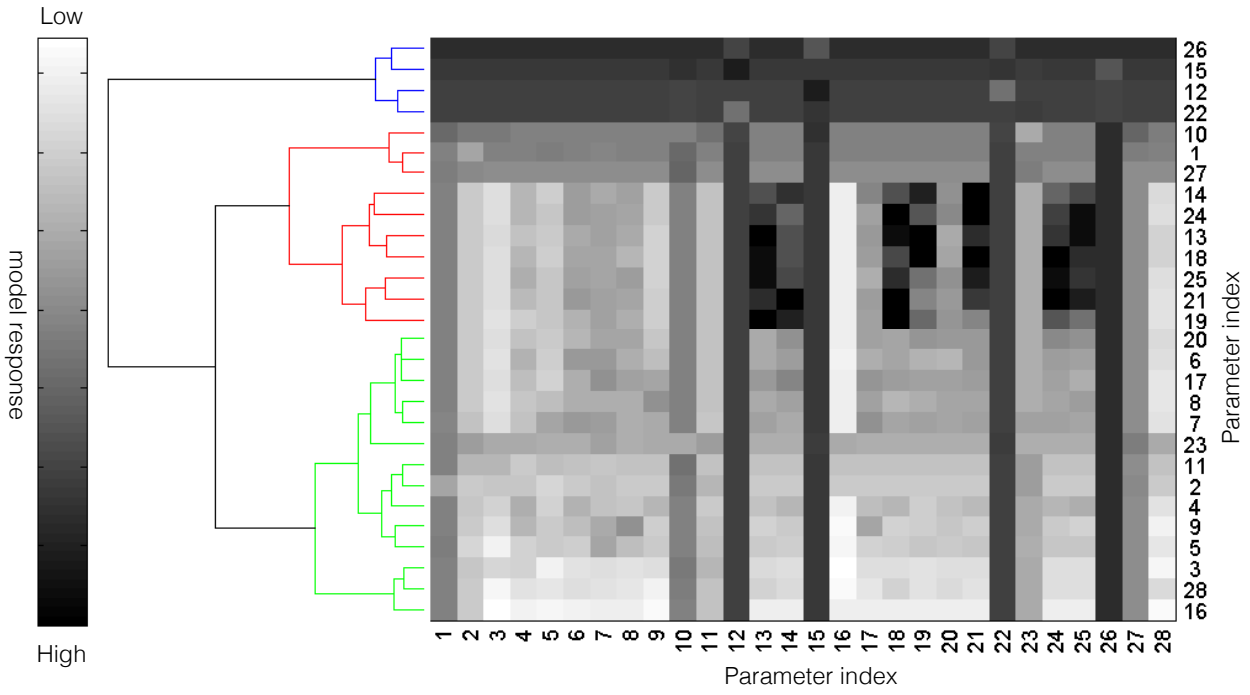

Figure 4.5: Pairwise sensitivity and clustering of complement model parameters in the presence of 1 mg/ml zymosan. The response of the complement model was calculated for each parameter combination following a 10% increase in parameter combinations in the presence of 1 mg/ml zymosan. The model parameters were clustered into high (blue), medium (red) and low (green) response clusters based upon the euclidian distance between the perturbed and nominal system state.

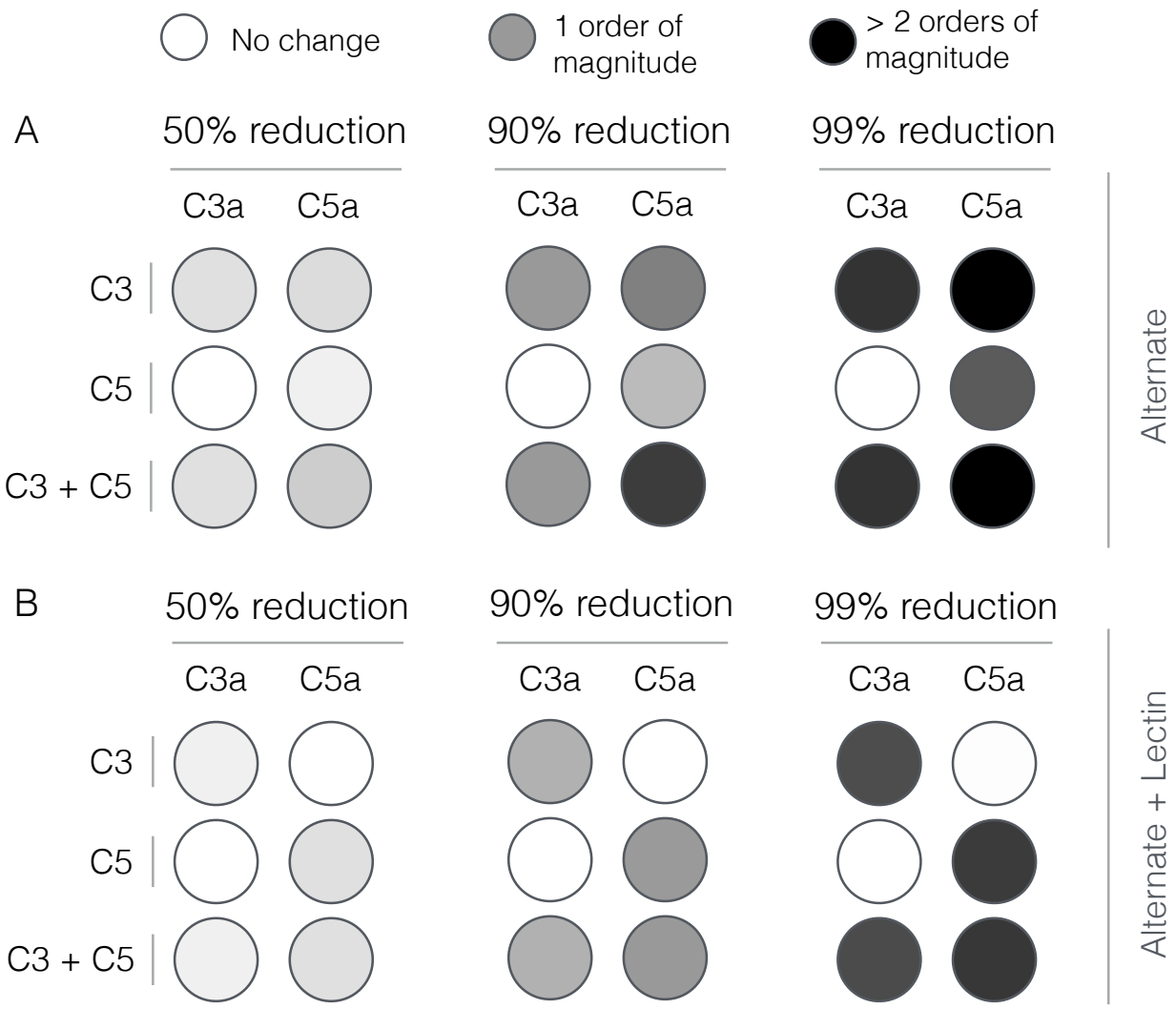

Figure 4.6: Robustness analysis of the complement model. Robustness coefficients were calculated for a 50%, 90% and 99% reduction in C3, C5, or C3 and C5 initial conditions. **A**: Mean robustness index for C3a and C5a generated from the alternate pathway (w/o zymosan). **B**: Mean robustness index for C3a and C5a generated from the lectin and alternate pathway (1 mg/ml zymosan). The color describes the degree of reduction of C3a or C5a following the network perturbation. Robustness coefficients were calculated using all parameter sets with Pareto rank less than five (N = 65). Mean robustness values were reported.

4.3 Discussion

In this study, we estimated an ensemble of experimentally validated reduced order complement models using multiobjective optimization. The modeling approach combined ordinary differential equations with logical rules to produce a complement model with a limited number of equations and parameters. The reduced order model, which described the lectin and alternative pathways, consisted of 18 differential equations with 28 parameters. Thus, the model was an order of magnitude smaller and included more pathways than comparable mathematical models in the literature. We estimated an ensemble of model parameters from *in vitro* time series measurements of the C3a and C5a complement proteins. Subsequently, we validated the model on unseen C3a and C5a measurements that were not used for model training. Despite its small size, the model was surprisingly predictive. After validation, we performed global sensitivity and robustness analysis to estimate which parameters and species controlled model performance. These analyses suggested complement was robust to any single therapeutic intervention. The only intervention that consistently reduced C3a and C5a formation for all cases was a knockdown of both C3 and C5. Taken together, we developed a reduced order complement model that was computationally inexpensive, and could easily be incorporated into pre-existing or new pharmacokinetic models of immune system function. The model described experimental data, and predicted the need for multiple points of intervention to disrupt complement activation.

Despite its importance, there has been a paucity of validated mathematical models of complement pathway activation. To our knowledge, this study is one of the first complement models that combined multiple initiation pathways with experimental validation of important complement products like C5a. However, there have been several theoret-

ical models of components of the cascade in the literature. Liu and co-workers modeled the formation of C3a through the classical pathway using 45 non-linear ODEs [149]. In contrast, in this study we modeled lectin mediated C3a formation using only five ODEs. Though we did not model all the initiation interactions in detail, especially the cross-talk between the lectin and classical pathways, we successfully captured C3a dynamics with respect to different concentrations of lectin initiators. The model also captured the dynamics of C3a and C5a formed from the alternate pathway using only seven ODEs. The reduced order model predictions of C5a were qualitatively similar to the theoretical complement model of Zewde et al., which involved over 100 ODEs [146]. However, we found that the C3a produced in the alternate pathway was nearly three orders of magnitude greater than the C5a generated. While this was in agreement with the experimental data [4], it differed from the theoretical predictions made by Zewde et al., who showed C3a was eight orders of magnitude greater than the C5a concentration [146]. In our model, the time profile of both C3a and C5a generated changed with respect to the quantity of zymosan (the lectin pathway initiator). In particular, the C3a peak time was directly proportional to initiator, while the lag phase for generation was inversely proportional to the initiator concentration. Korotaevskiy et al. showed a similar trend using a theoretical model of complement, albeit for much shorter time scales [148]. Thus, the reduced order complement model performed at least as well as existing larger mechanistic models, despite being significantly smaller.

Global analysis of the complement model suggested potentially important therapeutic targets. Complement malfunctions are implicated in a spectrum of diseases, however the development of complement specific therapeutics has been challenging [5, 133]. Previously, we have shown that mathematical modeling and analysis can be useful tools to estimate therapeutically important mechanisms [44, 63, 151, 152]. In this study, we an-

alyzed a validated ensemble of reduced order complement models to better understand the strengths and weaknesses of the cascade. In the presence of an initiator, C3a and C5a formation was sensitive to CP C3/C5 convertase assembly and activity, and to a lesser extent lectin initiation parameters. Formation of the CP convertases can be inhibited by targeting upstream protease complexes like MASP-1,2 from the lectin pathway (or C1r, C1s from classical pathway). For example, Omeros, a protease inhibitor that targets the MASP-2 complex, has been shown to inhibit the formation of downstream convertases [153]. Lampalizumab and Bikaciomab, which target factor B and factor D respectively, or naturally occurring proteins such as Cobra Venom Factor (CVF), an analogue of C3b, could also attenuate AP convertase formation [154–156]. Removing supporting molecules could also destabilize the convertases. For example, Novelmed Therapeutics developed the antibody, NM9401 against propedin, a small protein that stabilizes alternate C3 convertase [157]. Lastly, convertase catalytic activity could be attenuated using small molecule protease inhibitors. All of these approaches are consistent with the results of the sensitivity analysis. On the other hand, robustness analysis suggested C3a and C5a generation could only be significantly attenuated by modulating the free levels of C3 and C5. The most commonly used anti-complement drug Eculizumab, targets the C5 protein [5]. Several other antibodies targeting C5 are also being developed; for example, LFG316 targets C5 in Age-Related Macular Degeneration [158], while Mubodina is used to treat Atypical Hemolytic-Uremic Syndrome (aHUS) [159]. Other agents such as Coversin [160] or the aptamer Zimura [161] could also be used to knockdown C5. The peptide inhibitor Compstatin and its derivatives are promising approaches for the inhibition of C3 [162]. However, while the knockdown of C3 and C5 affect C3a and C5a levels downstream, the abundance, turnover rate and population variation of these proteins make them difficult targets [163, 164]. For example, the eculizumab dosage must be significantly adjusted during the course of treatment for aHUS [165]. A validated complement model, in com-

bination with personalized pharmacokinetic models of immune system function, could be an important development for the field.

The performance of the complement model was impressive given its limited size. However, there are several questions that should be explored further. A logical progression for this work would be to expand the network to include the classical pathway and the formation of the membrane attack complex (MAC). However, time course measurements of MAC abundance (and MAC formation dynamics) are scarce, making the inclusion of MAC challenging. On the other hand, inclusion of classical pathway activation is straightforward. Liu et al., have shown cross-talk between the activation of the classical and lectin pathways through C reactive proteins (CRP) and L-ficolin (LF) under inflammation conditions [149]. Thus, inclusion of these species, in addition to a lumped activation term for the classical pathway should allow us to capture classical activation. Next, we should address the C3a time scale issue. We believe the C3a time scale was related to our choice of training data, how we modeled the tickover mechanism, and factor B and D limitation. Tickover was modeled as a first-order generation processes where C3wBb formation and activity was lumped into the AP C3 convertase. Thus, we skipped an important step which could strongly influence AP C3 convertase formation by slowing down the rate C3 cleavage into C3a and C3b. The model should be expanded to include the C3wBb intermediate, where C3wBb catalyzes C3 cleavage at a slow rate compared to normal AP or CP C3 convertases. We also assumed both factor B and factor D were not limiting, thereby artificially accelerating the rate of AP C3 convertase formation. This shortcoming could be addressed by including balances around factor B and D, and including these species in the appropriate kinetic rates. The C5a predictions also had an accelerated time scale. However, because the C5a time scale depended strongly upon C3 convertase formation, we can likely correct the C5 issues by fixing the rate of C3 cleav-

age. Lastly, we should also consider including the C2-bypass pathway, which was not included in the model. The C2-bypass mediates lectin pathway activation, without the involvement of MASP-1/2. Thus, this pathway could be important for understanding the role of MASP-1/2 inhibitors on complement activation.

4.4 Materials and Methods

4.4.1 Formulation and solution of the complement model equations.

We used ordinary differential equations (ODEs) to model the time evolution of complement proteins (x_i) in the reduced order model:

$$\frac{1}{\tau_i}\frac{dx_i}{dt} = \sum_{j=1}^{\mathcal{R}} \sigma_{ij} r_j(\mathbf{x}, \epsilon, \mathbf{k}) \qquad i = 1, 2, \ldots, \mathcal{M} \tag{4.1}$$

where $\mathcal{R}$ denotes the number of reactions and $\mathcal{M}$ denotes the number of proteins in the model. The quantity τ_i denotes a time scale parameter for species i which captures unmodeled effects. For the current study, τ scaled with the level of initiator (z) for C5a and C5b; $\tau_i = z/z^*$ for i = C5a, C5b where z^* was 1mg/ml, $\tau_i = 1$ for all other species. The quantity $r_j(\mathbf{x}, \epsilon, \mathbf{k})$ denotes the rate of reaction j. Typically, reaction j is a non-linear function of biochemical and enzyme species abundance, as well as unknown model parameters $\mathbf{k}$ ($\mathcal{K} \times 1$). The quantity σ_{ij} denotes the stoichiometric coefficient for species i in reaction j. If $\sigma_{ij} > 0$, species i is produced by reaction j. Conversely, if $\sigma_{ij} < 0$, species i is consumed by reaction j, while $\sigma_{ij} = 0$ indicates species i is not connected with reaction j. Species balances were subject to the initial conditions $\mathbf{x}(t_o) = \mathbf{x}_o$.

Rate processes were written as the product of a kinetic term ($\bar{r}_j$) and a control term (v_j) in the complement model. The kinetic term for the formation of C4a, C4b, C2a and C2b, lectin pathway activation, and C3 and C5 convertase activity was given by:

$$\bar{r}_j = k_j^{max} \epsilon_i \left(\frac{x_s^\eta}{K_{js}^\eta + x_s^\eta} \right) \tag{4.2}$$

where k_j^{max} denotes the maximum rate for reaction j, ϵ_i denotes the abundance of the enzyme catalyzing reaction j, η denotes a cooperativity parameter, and K_{js} denotes the saturation constant for species s in reaction j. We used mass action kinetics to model protein-protein binding interactions within the network:

$$\bar{r}_j = k_j^{max} \prod_{s \in m_j^-} x_s^{-\sigma_{sj}} \tag{4.3}$$

where k_j^{max} denotes the maximum rate for reaction j, σ_{sj} denotes the stoichiometric coefficient for species s in reaction j, and $s \in m_j$ denotes the set of *reactants* for reaction j. We assumed all binding interactions were irreversible.

The control terms $0 \leq v_j \leq 1$ depended upon the combination of factors which influenced rate process j. For each rate, we used a rule-based approach to select from competing control factors. If rate j was influenced by $1, \ldots, m$ factors, we modeled this relationship as $v_j = \mathcal{I}_j\left(f_{1j}(\cdot), \ldots, f_{mj}(\cdot)\right)$ where $0 \leq f_{ij}(\cdot) \leq 1$ denotes a regulatory transfer function quantifying the influence of factor i on rate j. The function $\mathcal{I}_j(\cdot)$ is an integration rule which maps the output of regulatory transfer functions into a control variable. Each regulatory transfer function was modeled using a Hill function. In this study, we used $\mathcal{I}_j \in \{min, max\}$ [166]. If a process has no modifying factors, $v_j = 1$. The model equations were implemented in Julia and solved using the `CVODE` routine of the Sundials package [167, 168]. The model code and parameter ensemble is freely available under an MIT software license and can be downloaded from the Varnerlab website [169].

Complement activation under flow conditions.

We estimated the dynamics of complement activation under flow using a two-compartment model, with variable compartment volumes. We considered a main compartment (m), and a wound compartment (w), where complement was activated in the wound compartment by the addition of a pathogenic surface. In the main compartment, the balance for species i ($x_{m,i}$) was given by:

$$\frac{V_m}{\tau_i}\frac{dx_{m,i}}{dt} = \left(\sum_{j=1}^{\mathcal{R}} \sigma_{ij} r_j\left(\mathbf{x}_m, \epsilon, \mathbf{k}\right)\right) V_m - k_{m,w,i} x_{m,i} + k_{w,m,i} x_{w,i} - C x_{m,i} \tag{4.4}$$

where V_m denotes the volume of the main compartment, $k_{p,q,i}$ denotes transfer constant governing the transfer of species i from compartment p to compartment q, and C denotes the clearance constant from the main compartment. We assumed complement factors were synthesized in the main compartment. The balance governing species i in the wound compartment was given by:

$$\frac{V_w}{\tau_i}\frac{dx_{w,i}}{dt} = \left(\sum_{j=1}^{\mathcal{R}} \sigma_{ij} r_j\left(\mathbf{x}_w, \epsilon, \mathbf{k}\right)\right) V_w - k_{w,m,i} x_{w,i} + k_{m,w,i} x_{m,i} - B x_{w,i} \tag{4.5}$$

where V_w denotes the volume of the wound compartment, and B denotes the rate of blood loss from the wound compartment. Lastly, because of the volume loss through clearance in the main compartment, and bleeding from the wound compartment, the volumes of each compartment were dynamic modeled:

$$\frac{dV_m}{dt} = I_m + F_{w,m} - F_{m,w} - C \tag{4.6}$$

$$\frac{dV_w}{dt} = I_w + F_{m,w} - F_{w,m} - B \tag{4.7}$$

where I_m, I_w denote the rate of liquid input into the main and wound compartment.

4.4.2 Estimating complement model parameters.

We estimated a single initial parameter set using the Dynamic Optimization with Particle Swarms (DOPS) technique [170]. DOPS is a novel hybrid meta-heuristic which combines a multi-swarm particle swarm method with the dynamically dimensioned search approach of Shoemaker and colleagues [171]. DOPS minimized the squared residual between simulated and C3a and C5a measurements with and without zymosan as a single objective. The best fit set estimated by DOPS served as the starting point for multiobjective ensemble generation using Pareto Optimal Ensemble Technique in the Julia programming language (JuPOETs) [150]. JuPOETs is a multiobjective approach which integrates simulated annealing with Pareto optimality to estimate model ensembles on or near the optimal tradeoff surface between competing training objectives. JuPOETs minimized training objectives of the form:

$$O_j(\mathbf{k}) = \sum_{i=1}^{\mathcal{T}_j}\left(\hat{\mathcal{M}}_{ij} - \hat{y}_{ij}(\mathbf{k})\right)^2 + \left(\frac{\mathcal{M}'_{ij} - \max y_{ij}}{\mathcal{M}'_{ij}}\right)^2 \tag{4.8}$$

subject to the model equations, initial conditions and parameter bounds $\mathcal{L} \leq \mathbf{k} \leq \mathcal{U}$. The first term in the objective function measured the shape difference between the simulations and measurements. The symbol $\hat{\mathcal{M}}_{ij}$ denotes a scaled experimental observation (from training set j) while the symbol $\hat{y}_{ij}$ denotes the scaled simulation output (from training set j). The quantity i denotes the sampled time-index and $\mathcal{T}_j$ denotes the number of time

points for experiment j. The scaled measurement is given by:

$$\hat{\mathcal{M}}_{ij} = \frac{\mathcal{M}_{ij} - \min_i \mathcal{M}_{ij}}{\max_i \mathcal{M}_{ij} - \min_i \mathcal{M}_{ij}} \tag{4.9}$$

Under this scaling, the lowest measured concentration become zero while the highest equaled one, where a similar scaling was defined for the simulation output. The second-term in the objective function quantified the absolute error in the estimated concentration scale, where the absolute measured concentration (denoted by $\mathcal{M}'_{ij}$) was compared with the largest simulated value. In this study, we minimized two training objectives, the total C3a and C5a residual w/o zymosan (O_1) and the total C3a and C5a residual for 1 mg/ml zymosan (O_2). JuPOETs identified an ensemble of N = 2100 parameter sets which were used for model simulations and uncertainty quantification subsequently. JuPOETs is open source, available under an MIT software license. The JuPOETs source code is freely available from the JuPOETs GitHub repository [172]. The objective functions used in this study are available in the GitHub model repository [169].

The simulation and prediction performance of the complement model was measured using the Akaike information criterion (AIC) [173]. In this study, we implemented the AIC as:

$$AIC = 2N_p + N_m \ln\left(\frac{1}{\mathcal{M}} \sum_{\tau} (x_\tau - y_\tau)^2\right) \tag{4.10}$$

where N_p, N_m denotes the number of parameters, and the number of experimental measurements, respectively. The summation term in Eq. (5.12) denotes the residual between the model simulation (x) and experimental measurements (y), where the residual is normalized by the scale of the experimental data ($\mathcal{M}$). We compared the AIC for the model parameters estimated in this study, with a random parameter control generated to have a similar order of magnitude. The mean and standard deviation of the AIC was calculated

over the parameter ensemble and the random parameter control were reported in this study.

Pairwise sensitivity analysis and clustering.

We perturbed each pair of model parameters by 10% of their nominal value, and then calculated the euclidian distance between the perturbed and nominal system states for physiological conditions. We repeated this calculation for each member of the parameter ensemble, and calculated the mean differences between the perturbed and nominal states. We then clustered the resulting `log10` transformed mean distances using the `Clustergram` routine in MATLAB (The Mathworks, Natick MA). We considered three clusters, high, medium and low displacement.

Robustness analysis.

Robustness coefficients quantify the response of a marker to a structural or operational perturbation to the network architecture. Robustness coefficients were calculated as shown previously [26]. Log-transformed robustness coefficients denoted by $\hat{\alpha}\left(i, j, t_o, t_f\right)$ were defined as:

$$\hat{\alpha}\left(i, j, t_o, t_f\right) = \log_{10}\left[\left(\int_{t_o}^{t_f} x_i\left(t\right) dt\right)^{-1}\left(\int_{t_o}^{t_f} x_i^{(j)}\left(t\right) dt\right)\right] \tag{4.11}$$

Here, t_o and t_f denote the initial and final simulation time, while i and j denote the indices for the marker and the perturbation, respectively. A value of $\hat{\alpha}\left(i, j, t_o, t_f\right) > 0$, indicates increased marker abundance, while $\hat{\alpha}\left(i, j, t_o, t_f\right) < 0$ indicates decreased marker abundance

following perturbation j. If $\hat{\alpha}\left(i, j, t_o, t_f\right) \sim 0$, perturbation j did not influence the abundance of marker i. In this study, we perturbed the initial condition of C3 or C5 or a combination of C3 and C5 by 50%, 90% and 99% and measured the area under the curve (AUC) of C3a or C5a with and without lectin initiator. We computed the robustness coefficients for a subset of the parameter ensemble (N = 65) and reported the mean robustness value.

CHAPTER 5

EFFECTIVE MODELING OF THE HUMAN COAGULATION AND FIBRINOLYTIC PATHWAYS

5.1 Introduction

Trauma [1] is the leading cause of death and disability, surpassing all other causes combined, for persons 36 years old and younger [174], accounting for 30% of the years of life lost in the United States at an economic burden of $671 billion a year, in health care costs and lost productivity [175]. Hemorrhage accounts for 40% of all trauma deaths, where the control of bleeding is especially challenging in the presence of blood coagulation disorders, collectively known as coagulopathy [176]. However, adverse outcomes associated with coagulopathy are not limited to death from acute blood loss. Organ dysfunction, multiple organ failure and increased susceptibility to sepsis [177] are all potential consequences of prolonged shock resulting from coagulopathy [178]. Following a wound, the immediate response of the body is to activate the coagulation cascade, which in turn generates a clot through the fibrinolysis network.

The coagulation cascade, which is activated following a wound, is mediated by a family proteases in the circulation, called factors and a key group of blood cells, called platelets. The central process in coagulation is the conversion of prothrombin (fII), an inactive coagulation factor, to the master protease thrombin (FIIa). Thrombin generation involves three phases, initiation, amplification and termination [34, 35]. Initiation requires a trigger event, for example vessel injury, which leads to the activation of factor VII (FVIIa).

[1] Adapted with permission from Sagar A, LeCover R, Brummel-Ziedins KE, Orfeo T, Varner JD et al. (2017) "Dynamic Modeling of Fibrinolysis using Reduced Order Effective Kinetic Models" *In preparation*

Two converging pathways, the extrinsic and intrinsic cascades, then process and amplify this initial coagulation signal. The extrinsic cascade is generally believed to be the main mechanism of thrombinogenesis in the blood [36–38]. Initially, thrombin is produced upon cleavage of prothrombin by fluid phase activated factor X (FXa), which itself has been activated by TF/FVIIa [31]. Picomolar amounts of thrombin then activate the cofactors factors V and VIII (fV and fVIII) and platelets, leading to the formation of the tenase and prothrombinase complexes on activated platelets. These complexes amplify the early coagulation signal by further activating FXa, and directly converting prothrombin to thrombin. There are several control points in the cascade that inhibit thrombin formation, and eventually terminate thrombin generation. Tissue Factor Pathway Inhibitor (TFPI) inhibits FXa formation catalyzed by TF/FVIIa, while antithrombin III (ATIII) neutralizes several of the proteases generated during coagulation, including thrombin. Thrombin itself also inadvertently plays a role in its own inhibition; thrombin, through interaction with thrombomodulin, protein C and endothelial cell protein C receptor (EPCR), converts protein C to activated protein C (APC) which attenuates the coagulation response by proteolytic cleavage of fV/FVa and fVIII/FVIIIa. Termination occurs after either prothrombin is consumed, or thrombin formation is neutralized by inhibitors such as APC or ATIII. Activated thrombin plays a dual role in blood clot formation. First, it catalyzes the cleavage of fibrinogen to fibrin, a key component of a blood clot, thereby promoting clot formation. Activated thrombin also activates thrombin-activatable fibrinolysis inhibitor (TAFI), which inhibits the activity of plasmin, an important enzyme present in blood that degrades many blood plasma proteins, including fibrin clots. Counter balancing the role of activated thrombin is tissue plasminogen activator (tPA), which activates plasmin, thereby promoting the break down of blood clots. Thus, a delicate balance exists between activated thrombin and tPA that controls the rate of clot formation. Too much activated thrombin leads to hypo-fibrinolysis (excessive clot formation resulting in stroke

or heart attack risk), while too little or excess tPA leads to hyper-fibrinolysis (inadequate clot formation resulting in increased, sometimes catastrophic bleeding). There are excellent reviews of both coagulation and fibrinolysis, see [67, 179].

We have worked to understand the coagulation cascade, using both mechanistic models e.g., [44, 180], and more recently reduced order modeling approaches [166]. Longstaff and Thelwell proposed a very simple model for fibrinolysis, in which the process was represented as two steps, with plasminogen being converted to plasmin by tPA, and fibrin degraded by plasmin [181]. While computationally easy to evauluate, this model greatly oversimplifies the process, and completely neglects the role of PAI-1. Fibrinolysis has been previous modeled in great detail through a 3D stochastic mutltiscale model which predicted difference in lysis speeds based on clot morphology, but this model focuses only clot lysis rather than formation and degredation [182]. Reifman et al modeled fibrin generation with mass action kinetics using 80 ODEs to investigate the effacy of different prothombin complex concentrates [79]. They used an interesting method to model the complex kinetics of TAFI inhibiting fibrinolysis: they emperically fit a curve to the clot lysis time as a function of TAFI concentration, and then correspondingly reduced the tPA effective concentration to account for the effects of TAFI. While this model decently predicted the final fibrin levels, it did a poor job of capturing the shape of the fibrin generation curves.

In this study, we developed a reduced order of model of coagulation and fibrinolysis. We created a function to transform the biochemical species concentrations predicted by the model into a ROTEM signal. We trained the model on ROTEM curves at various concentrations of tPA, and then successfully predicted previously unseen data. We used a variance based method to identify the most sensitive parameters. We then clustered the parameters using a hierarichal clustering method to determine the interactions be-

tween parameters. We found that our reduced order model was successful in predicting ROTEM curves. We believe that this is the first study to connect a kinetic model with ROTEM curves, a key development, as ROTEM is playing an increasingly large role both in surgery and in assessing the coagulation state of a patient. Future refinements include adding platelets to the model as activated platelets greatly influence the activation rates of prothrombin and FX [183].

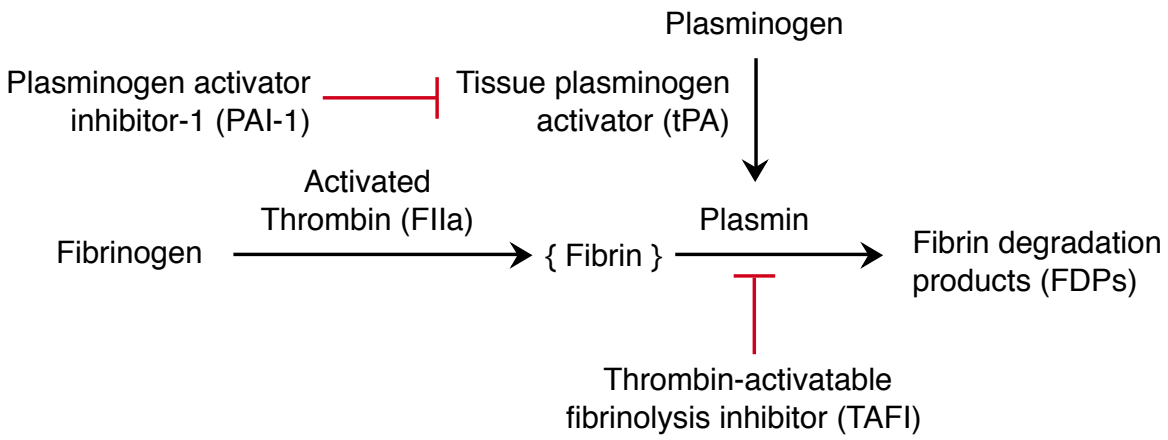

Figure 5.1: Schematic of the fibrinolysis model.

5.2 Results

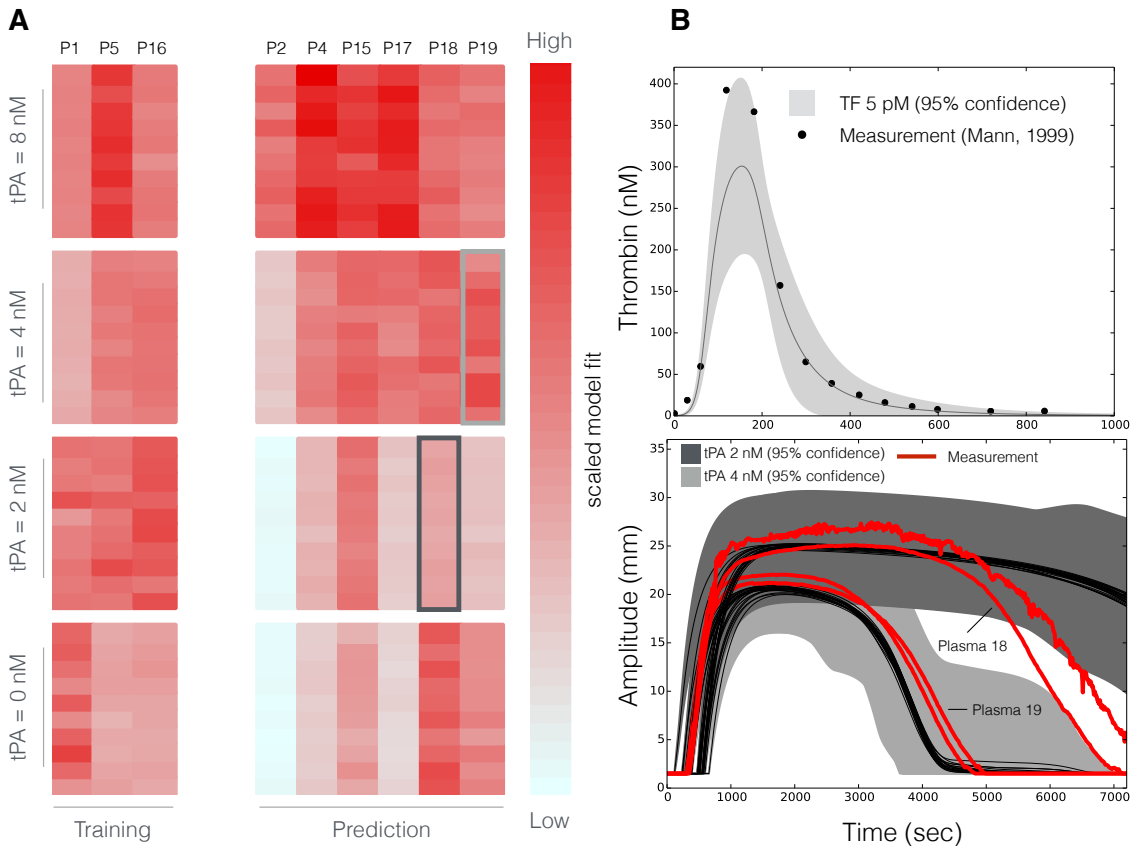

Figure 5.2: Fibrinolysis and coagulation model training and prediction.

5.2.1 Sensitivity analysis.

We calculated total order sensitivity coefficients by the method of Sobol with respect to the area under the ROTEM curve to understand which parameter changes would have the most dramatic effect on model output (5.3). We found that among the kinetic parameters, the rate constants for fibrin and protofibril association had a large effect, as well as the rate constant for plasmin/anti-plasmin association. The turnover number for tPA was the most sensitive parameter, and the turnover number for plasmin also was found to be

sensitive. The majority of the control parameters were found to be fairly non-sensitive, with the exception of the parameters controlling TAFI. When we considered initial conditions, we found that the model was much more sensitive to the starting concentration of tPA than the initial concentrations of plasminogen or TAFI. We recalculated the total order sensitivity coefficients for a patient experiencing severe hemodilution and found that the model was still highly sensitive to changes in the initial concentration of tPA, but had decreased slightly in sensitivity to the concentration of plasminogen. In this hemodiluted case, the sensitivities of the kinetic parameters remained nearly unaltered, but sensitivity of the TAFI control parameters decreased. While the method of Sobol was useful in explaining changes in the area under of the curve, we also sought to examine the changes to the shape of the ROTEM curve.

We used the method of Morris, combined with functional principle component analysis to examine how changes in parameter values altered the ROTEM curve (5.4). This method found that the control parameters related to thrombin generation and inhibition, as well as the rate constant for antithrombin inhibiting activation of thrombin to have large effects on the curve shape, as signified by large values of the estimate of the mean of the distribution of the absolute values of the elementary effects. The estimate of these parameters' elementary effects had large standard deviations, indicating that they effect the model output in a non-linear way. All of these parameters determine the amount of thrombin available, so it is logical that changes in their values would dramatically alter the ROTEM curve. While the turnover number was found to be very sensitive by the method of Sobol in changing the area under the ROTEM curve, the method of Morris found it to only have a moderate effect on the shape of the ROTEM curve. While both of these techniques informed us about how single parameter changes affected the model output, we wished to investigate interactions between multiple parameters, as it is very

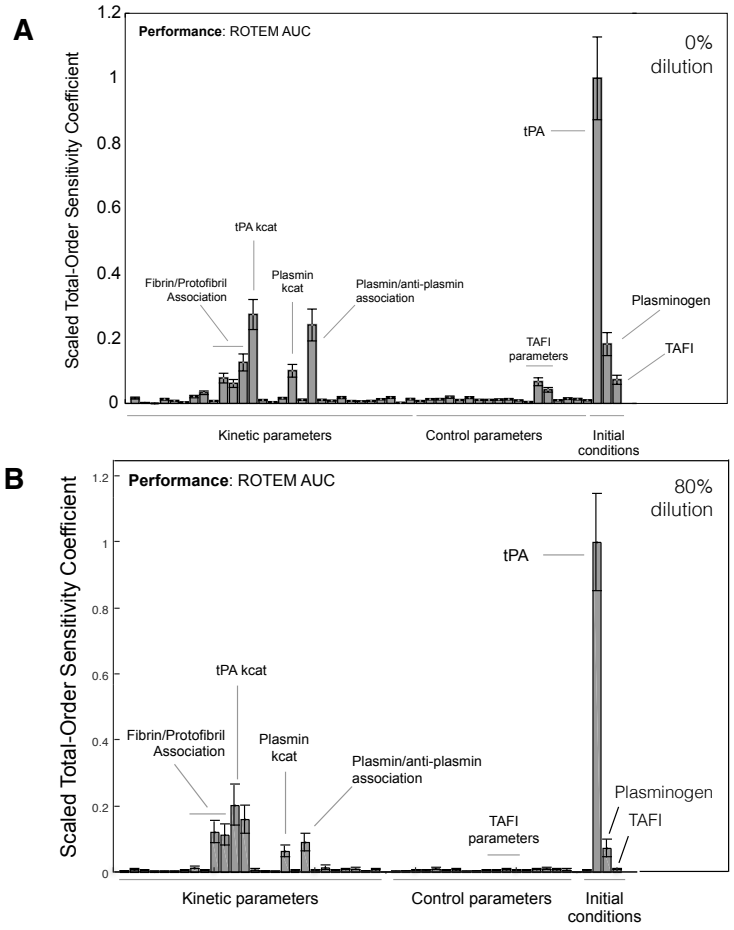

Figure 5.3: Global sensitivity analysis of the fibrinolysis model.

rare in biological systems for only one element to change while the rest are held perfectly constant.

We used hierarchical clustering to examine the effects of pairwise perturbing the model parameters (5.5). This resulted in three clusters of parameters, each shown in a different color with parameters which had the largest impact on the ROTEM curve in green, moderate impact in blue, and small impact in red. The parameters in the cluster with the largest impact are all parameters related to clot formation and dissolution, with

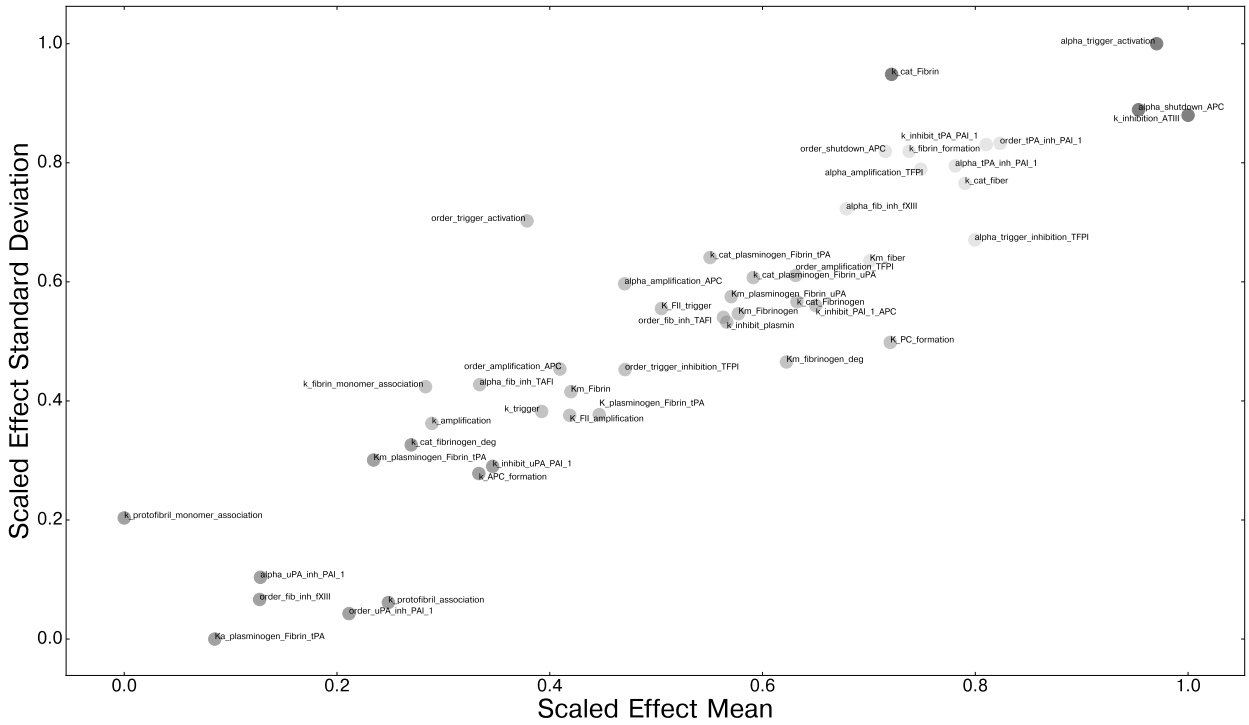

Figure 5.4: Morris analysis of fibrinolysis model response.

the exception of the rate constant for ATIII inhibiting thrombin. Changes in these parameters in combination with nearly any other parameter in the model resulted in significant changes in the ROTEM curve, suggesting that they may be robust therapeutic targets. The moderate impact cluster includes rate constants for plasminogen activation as well as clot formation, thrombin formation, and initiation. There is one interaction of interest in this cluster: a proportionate increase in the rate constant for amplification cancels out a change in the control parameters for TFPI. This perhaps arises because both parameters have opposing effects on thrombin generation. The cluster containing parameters with the smallest impact contains a larger number of control parameters, as well as the rate constants for the interactions between uPA and plasminogen. Among all the control parameters used in the model, only three were found to have a moderate or high impact:

those controlling either FXIII or TAFI. We then wished to observe the effects of the initial conditions on the model output.

We again used hierarchical clustering to observe the results of pairwise altering the model initial conditions (5.6). We found that doubling the initial concentration of tPA, in combination with increasing any other initial condition, resulted in a large change in the predicted ROTEM curve. Plasminogen and TAFI, which both had moderately sized total order sensitivity coefficients, as estimated by the method of Sobol, both had different effects, which combined with other changes in initial conditions. The initial amount of fibrinogen present also played a significant role in the model's output, especially when changes in the amount of fibrinogen were combined with a change in tPA levels. However, altering the concentration of many of the enzymes involved in the coagulation cascade (FV, FVIII, FX, FIX) did not result in large changes, unless paired with a change in another species which had a powerful effect on its own, such as tPA or fibrinogen.

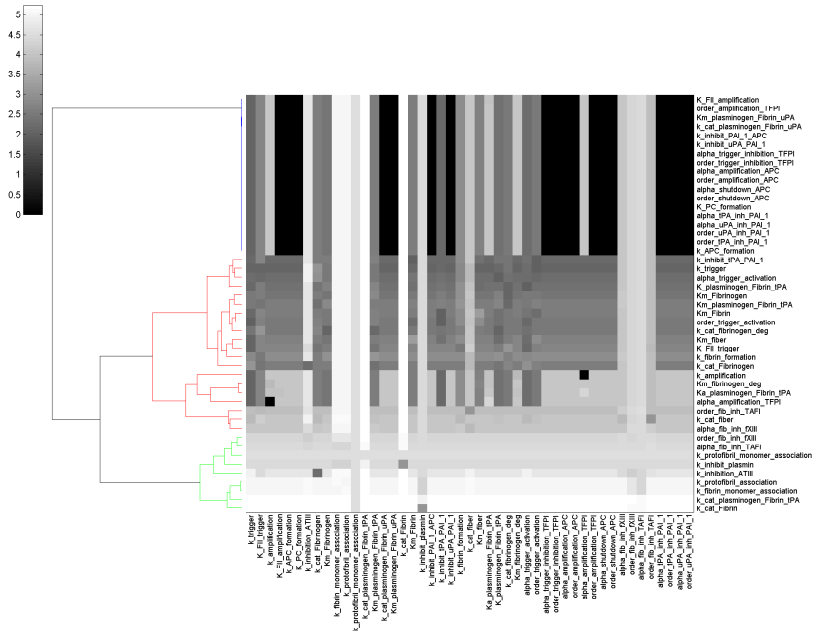

Figure 5.5: Clustergram analysis of fibrinolysis model response to changes in model parameters.

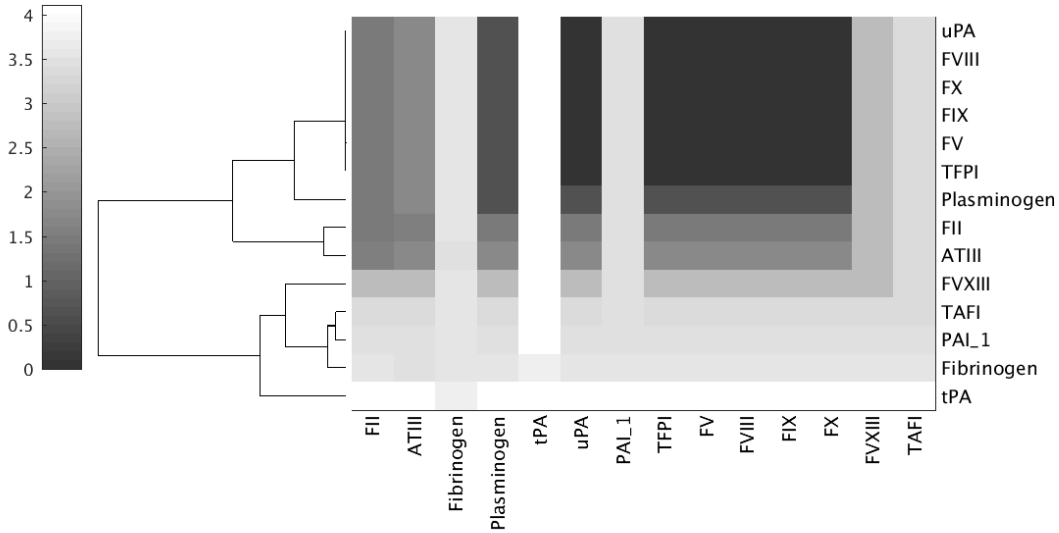

Figure 5.6: Clustergram analysis of fibrinolysis model response to changes in initial conditions.

5.3 Discussion

In this study, we constructed a reduced order model of fibrinolysis, which considered the dynamics of thrombin activation. We then created a function to transform the concentrations into ROTEM curves, and estimated kinetic and control parameters for the models using a Pareto Optimal Ensemble Technique. We then validated the model by measuring its performance on datasets that were not used in the parameter estimation. We then performed sensitivity analysis to observe which targets in the system could potentially be therapeutic targets.

Both the total order sensitivity constants and the hierarchical clustering identify tPA's k_{cat} as sensitive, even under dilution conditions. Therefore, it could potentially serve as a therapeutic target. Additionally, the initial condition of tPA was very sensitive in both cases, suggesting that if a drug could render tPA inactive or increase its concentration, clot breakdown times could be significantly altered. The drug Alteplase (Cathflow Activase) does exactly that, given that it consists of recombinant version of tPA [184]. The method of Morris found that the rate constant of ATIII inhibiting thrombin to have a large effect on model output, which is consistent with heparin causing a conformation change in ATIII and altering its catalytic properties so that the reaction between it and thrombin goes much faster [185].

In the present model, the formation of a clot is represented by the formation of fiber, which is formed by protofibrils. The polymerization of protofibrils into oligomers, and the growth of the oligomers into a clot could potentially be included to further refine the model. Furthermore, the rate at which plasmin breaks down a clot is a function of the clot's structure [186]. Future work could include generating different types of fibers, with differing kinetic properties with respect to plasmin. Additionally, the role of platelets was

completely neglected in the clot formation process. It is challenging to include platelets in a kinetic model, as their concentration is many orders of magnitude smaller than the proteins involved in coagulation and fibrinolysis (femtomolar vs nanomolar), so instead of directly modeling the kinetics of activation, we can perhaps model the fraction of platelets activated, and use that to modify the control terms.

5.4 Materials and Methods

5.4.1 Materials.

Full-length (residues 1-263) recombinant tissue factor (Tf) was purchased from Haematologic Technologies (Essex Junction VT, USA). 1,2-Dioleolyl-sn-Glycero-3-Phospho-L-Serine (PS) and 1,2-Dioleoyl-sn-Glycero-3-Phosphocholine (PC) were purchased from Avanti Polar Lipids, Inc. (Alabaster AL, USA). Preparations of phospholipid vesicles (PCPS) composed of 75% PC and 25% PS were made as described [187] as was the TF/PCPS reagent [188]. Corn trypsin inhibitor (CTI) was purified in-house [188] as was D-Phe-Pro-Arg-CH2Cl (FPR-ck) [189]. Tissue plasminogen activator (t-PA) (two chain) was purchased from Molecular Innovations, Inc., MI, USA). α2-antiplasmin was purchased from Enzyme Research Labs Inc. (IN, USA) and plasmin from Haematologic Technologies Inc. (Essex Junction VT,USA). Aprotinin was purchased from Sekisui Diagnostics (Stamford CT, USA). Units of transfusion plasma were obtained from the University of Vermont Medical Center Blood Bank Plasma from each unit was distributed into aliquots and immediately frozen at -80^{o}C until used.

5.4.2 Assays.

Assays assessing the functional level of fibrinogen were performed by the University of Vermont Medical Center Clinical Laboratory. ELISA methods were used to assess protein antigen levels: plasminogen (PG-EIA, Affinity Biologicals ON, Canada); thrombin activatable fibrinolysis inhibitor (TAFI) (Zymutest (TAFI) Total Ag, Hyphen Biomedical, Nevville sur Olse, FR); plasminogen activator inhibitor 1 (PAI-1) (R&D Systems Inc., MN,

USA). α-thrombin-antithrombin complex (α-TAT) levels were determined via in house ELISA as described previously [190]. Plasmin-antiplasmin (PAP) complex levels were determined using the Imunoclone PAP ELISA (Sekisui Diagnostics CT, USA) with the following modification. A stock of 6 μM PAP complex was generated in house by reacting 9 μM α2 antiplasmin (Enzyme Research Labs Inc., IN, USA) with 6 μM plasmin (Haematologic Technologies Inc., VT,USA). Quantitative formation of PAP was confirmed via SDS-PAGE analysis. This stock was used in place of the kit calibrator to generate a standard curve.

5.4.3 Viscoelastometry.

Citrate plasmas were recalcified (15 mM CaCl2 final, volume change of 1.5%) for 3 min at 37^{o}C. The tissue factor reagent (20 μL, 5 pM final) and t-PA (20 μL, 2 nM, 4 nM or 8 nM final) or 20 μL buffer (no t-PA control) were prealiquoted as separate drops into 8 ROTEM cups. Each reaction was run in duplicate. 300 μL of recalcified plasma was then added to each cup and data collection initiated. The contents of ROTEM cups were collected as described previously [190, 191].

5.4.4 Fibrinolysis model equations.

We used ordinary differential equations (ODEs) to model the time evolution of fibrinolysis and coagulation proteins (x_i) in the reduced order model:

$$\frac{dx_i}{dt} = \sum_{j=1}^{\mathcal{R}} \sigma_{ij} r_j(\mathbf{x}, \epsilon, \mathbf{k}) \qquad i = 1, 2, \ldots, \mathcal{M} \tag{5.1}$$

$$\mathcal{A}(\mathbf{x}) = \mathcal{A}_o + \mathcal{A}_1(\mathbf{x}) \left[\frac{\mathcal{F}(\mathbf{x})^2}{K(\mathbf{x})^2 + \mathcal{F}(\mathbf{x})^2} \right] \tag{5.2}$$

where $\mathcal{R}$ denotes the number of reactions, $\mathcal{M}$ denotes the number of proteins in the model, x_i denotes the concentration of species i, and $\mathcal{A}$ denotes the output function which transforms biochemical model species into a ROTEM signal. The quantity $r_j(\mathbf{x}, \epsilon, \mathbf{k})$ denotes the rate of reaction j. Typically, reaction j is a non-linear function of biochemical ($\mathbf{x}$) and enzyme (ϵ) species abundance, as well as unknown model parameters $\mathbf{k}$ ($\mathcal{K} \times 1$). The quantity σ_{ij} denotes the stoichiometric coefficient for species i in reaction j. If $\sigma_{ij} > 0$, species i is produced by reaction j. Conversely, if $\sigma_{ij} < 0$, species i is consumed by reaction j, while $\sigma_{ij} = 0$ indicates species i is not connected with reaction j. Species balances were subject to the initial conditions $\mathbf{x}(t_o) = \mathbf{x}_o$.

Rate processes were written as the product of a kinetic term ($\bar{r}_j$) and a control term (v_j). Enzyme catalyzed rates were modeled using multiple saturation kinetics:

$$\bar{r}_j = k_j^{max} \epsilon_i \prod_{s \in m_j^-} \left(\frac{x_s}{K_{js} + x_s} \right) \tag{5.3}$$

where k_j^{max} denotes the maximum rate for reaction j, ϵ_i denotes the abundance of the enzyme catalyzing reaction j, K_{js} denotes the saturation constant for species s in reaction

j, and $s \in m_j^-$. We used mass action kinetics to model protein-protein binding interactions within the network:

$$\bar{r}_j = k_j^{max} \prod_{s \in m_j^-} x_s^{-\sigma_{sj}} \tag{5.4}$$

where k_j^{max} denotes the maximum rate for reaction j, σ_{sj} denotes the stoichiometric coefficient for species s in reaction j, and $s \in m_j$ denotes the set of *reactants* for reaction j. We assumed all binding interactions were irreversible.

The control terms $0 \leq v_j \leq 1$ depended upon the combination of factors which influenced rate process j. For each rate, we used a rule-based approach to select from competing control factors. If rate j was influenced by $1, \ldots, m$ factors, we modeled this relationship as $v_j = \mathcal{I}_j\left(f_{1j}(\cdot), \ldots, f_{mj}(\cdot)\right)$ where $0 \leq f_{ij}(\cdot) \leq 1$ denotes a regulatory transfer function quantifying the influence of factor i on rate j. The function $\mathcal{I}_j(\cdot)$ is an integration rule which maps the output of regulatory transfer functions into a control variable. In this study, we used $\mathcal{I}_j \in \{min, max\}$ and hill-like transfer functions [166]. If a rate process had no modifying factors, $v_j = 1$.

The output function $\mathcal{A}(\mathbf{x})$ transforms the biochemical species abundance predicted by the model into a ROTEM signal. Modeling the connection between blood rheology and the local fibrin concentration is complex and beyond the scope of this study. Instead, we developed an empirical model which consisted of a scale function, which captured the amplitude of the ROTEM signal, and a shape function which captured the signal shape. The quantity $\mathcal{A}_o$ denotes the baseline ROTEM measurement which we assumed was constant across plasma samples ($\mathcal{A}_o \simeq 1.5$). On the other hand, the shape function:

$$0 \leq \frac{\mathcal{F}(\mathbf{x})^2}{K(\mathbf{x})^2 + \mathcal{F}(\mathbf{x})^2} \leq 1 \tag{5.5}$$

was modified by a scale function. The quantity $\mathcal{F}(\mathbf{x})$ denotes the unweighted sum of fibrin species in the model, while the saturation function $K(\mathbf{x})$ was given by:

$$K(tPA) = 5000 - 375 \times (tPA) \tag{5.6}$$

The scale function $\mathcal{A}_1(\mathbf{x})$ was given by:

$$\mathcal{A}_1(\mathbf{x}) = \begin{cases} t \leq t_d & 0 \\ t > t_d & \mathcal{S} \times [1 - \exp(-\tau\,(t - t_d))] \end{cases} \tag{5.7}$$

where

$$\tau = 0.0035 \times \left[1 - \frac{FIIa}{\max(FII)}\right] \tag{5.8}$$

The material model delay parameter, $t_d = \mathcal{N}(\mu_d, \sigma_d)$, was identified from the training data. The scale parameter $\mathcal{S}$, modeled as a linear function of tPA abundance, was defined as:

$$\mathcal{S}(tPA) = \mathcal{N}(\mu_1, \sigma_1) - \mathcal{N}(\mu_2, \sigma_2) \times (tPA) \tag{5.9}$$

where the intercept and slope distributions were modeled as Gaussian distributions, where the mean and standard deviation for each distribution (μ_i, σ_i) were estimated directly from the training data. The model and output equations were implemented in the Julia programming language and solved using the `ODE23s` routine of the `ODE` package [167]. The model code and parameter ensemble is freely available under an MIT software license and can be downloaded from http://www.varnerlab.org.

5.4.5 Estimating fibrinolysis model parameters.

We estimated an ensemble of model parameters using Pareto Optimal Ensemble Technique in the Julia programming language (JuPOETs) [150]. JuPOETs is a multiobjective approach which integrates simulated annealing with Pareto optimality to estimate model ensembles on or near the optimal tradeoff surface between competing training objectives. JuPOETs minimized training objectives of the form:

$$O_j(\mathbf{k}) = \sum_{i=1}^{\mathcal{T}_j}\left(\hat{\mathcal{M}}_{ij} - \hat{y}_{ij}(\mathbf{k})\right)^2 + \left(\frac{\mathcal{M}'_{ij} - \max y_{ij}}{\mathcal{M}'_{ij}}\right)^2 \tag{5.10}$$

subject to the model equations, initial conditions and parameter bounds $\mathcal{L} \leq \mathbf{k} \leq \mathcal{U}$. The first term in the objective function measured the shape difference between the simulations and measurements. The symbol $\hat{\mathcal{M}}_{ij}$ denotes a scaled experimental observation (from training set j) while the symbol $\hat{y}_{ij}$ denotes the scaled simulation output (from training set j). The quantity i denotes the sampled time-index and $\mathcal{T}_j$ denotes the number of time points for experiment j. The scaled measurement is given by:

$$\hat{\mathcal{M}}_{ij} = \frac{\mathcal{M}_{ij} - \min_i \mathcal{M}_{ij}}{\max_i \mathcal{M}_{ij} - \min_i \mathcal{M}_{ij}} \tag{5.11}$$

Under this scaling, the lowest measured concentration become zero while the highest equaled one, where a similar scaling was defined for the simulation output. The second-term in the objective function quantified the absolute error in the estimated concentration scale, where the absolute measured concentration (denoted by $\mathcal{M}'_{ij}$) was compared with the largest simulated value. In this study, we minimized three training objectives, the total error estimated for tPA = 0, 2, 4, 8 nM for three plasmas from our plasma data set. JuPOETs identified an ensemble of N $\simeq$ 2500 parameter sets which were used for model

simulations and uncertainty quantification subsequently. JuPOETs is open source, available under an MIT software license. The JuPOETs source code is freely available from the JuPOETs GitHub repository at https://github.com/varnerlab/POETs.jl. The objective functions used in this study are available in the GitHub model repository available from http://varnerlab.org.

The simulation and prediction performance of the fibrinolysis model was measured using the Akaike information criterion (AIC) [173]. In this study, we implemented the AIC as:

$$AIC = 2N_p + N_m \ln\left(\frac{1}{\mathcal{M}}\sum_{\tau}(x_\tau - y_\tau)^2\right) \tag{5.12}$$

where N_p, N_m denotes the number of parameters, and the number of experimental measurements, respectively. The summation term in Eq. (5.12) denotes the residual between the model simulation (x) and experimental measurements (y), where the residual is normalized by the scale of the experimental data ($\mathcal{M}$). We compared the AIC for the model parameters estimated in this study, with a random parameter control generated to have a similar order of magnitude. The mean and standard deviation of the AIC was calculated over the parameter ensemble reported in this study.

5.4.6 Sensitivity analysis.

We conducted global sensitivity analysis to estimate which parameters controlled the performance of the reduced order model. We computed the total variance-based sensitivity index of each parameter relative to the training residual for the C3a/C5a alternate and C3a/C5a lectin objectives using the Sobol method [85]. sThe sampling bounds for each parameter were established from the minimum and maximum value for that parameter

in the parameter ensemble. We used the sampling method of Saltelli *et al.* to compute a family of $N(2d + 2)$ parameter sets which obeyed our parameter ranges, where N was the number of trials per parameters, and d was the number of parameters in the model [192]. In our case, $N = 400$ and $d = 28$, so the total sensitivity indices were computed using 23,200 model evaluations. The variance-based sensitivity analysis was conducted using the SALib module encoded in the Python programming language [87].

5.4.7 Morris Sensitivity Analysis

We performed global sensitivity analysis to determine which parameters controlled the shape of the ROTEM curve output by the model. We generated parameters using the Morris method, with the parameters bounded by the minimum and maximum values they took in the rank one parameter sets generated by JuPOETS [193]. A total of $N(p + 1)$ parameter sets were generated, with $N = 1000$ and $p = 47$ in our case, for a total of 48,000 model evaluations.We ran the model with the generated parameter sets, and then transformed the ROTEM curves using fPCA (functional principle component analysis), as implemented by the R package fdapace, to extract the principle component scores. This technique allowed us to remove the time dimension and to observe how the changes in parameters effected the shape of the ROTEM curve [194]. We then used the principle component scores to calculate μ^*, the absolute value of the mean of the elementary effect, and σ, the variance of the elementary effect. We opted to use the absolute values of the means of the elementary effects so that effects with opposite signs would not cancel each other out and artificially decrease the estimate of the mean of the elementary effect [195]. We then scaled μ^* and σ so that the smallest value became zero and the largest value became one, using the same scaling as used for the scaled experimental measurements.

This method sensitivity analysis was conducted using the SALib module encoded in the Python programming language [87]. We then used k-means clustering to group the parameters into four clusters, and colored the graph by cluster color. We used four clusters as it resulted in the best average silhouette score [196].

5.4.8 Clustering.

We perturbed each pair of model parameters to five times their nominal value, and then calculated the euclidean distance between the perturbed and nominal curves for physiological conditions. We clustered the `log10` transformed distances using the `Clustergram` routine in Matlab (The Mathworks, Natick MA). For the initial conditions, we perturbed them to one half and two times their nominal values in a pair wise fashion. We clustered the `log10` distances with a linear shift to make all of the values positive.

CHAPTER 6

DYNAMIC OPTIMIZATION WITH PARTICLE SWARMS (DOPS): A META-HEURISTIC FOR PARAMETER ESTIMATION IN BIOCHEMICAL MODELS

Abstract

Mathematical modeling is a powerful tool to analyze, and design biochemical networks. However, the estimation of biochemical model parameters is a significant challenge. Parameter estimation typically involves expensive function evaluations and noisy data, making it difficult to quickly obtain optimal solutions. Biochemical models often also have many local extrema which further complicates parameter estimation. Toward these challenges, we developed Dynamic Optimization with Particle Swarms (DOPS), a novel hybrid meta-heuristic that combined multi-swarm particle swarm optimization with dynamically dimensioned search (DDS) [1]. DOPS uses a multi-swarm particle swarm optimization technique to generate candidate solution vectors, the best of which is greedily updated using dynamically dimensioned search. We tested DOPS using a combination of classic optimization test functions, biochemical benchmark problems and real-world biochemical models. We performed $\mathcal{T} = 25$ trials with $N = 4000$ function evaluations per trial, and compared the performance of DOPS with other commonly used meta-heuristics such as differential evolution (DE), simulated annealing (SA) and dynamically dimensioned search (DDS). On average, DOPS outperformed other common meta-heuristics on the optimization test functions, benchmark problems and a model of the human coagulation cascade. Taken together, DOPS is a promising meta-heuristic approach for the estimation

[1] Adapted with permission from Sagar A, Shoemaker CA, Varner J (2016) "Dynamic Optimization with Particle Swarms (DOPS): A meta- heuristic for parameter estimation in biochemical models" Biotechnol J *In review*

of biochemical model parameters in relatively few function evaluations. DOPS source code is available for download under a MIT license at http://www.varnerlab.org.

6.1 Introduction

Cells process nutrients and respond to changes in their environment using complex enzyme catalyzed biochemical networks. Mathematical modeling has evolved as a powerful paradigm to analyze, and ultimately design these complex networks [197–201]. Mathematical modeling of biochemical networks is often an iterative process. First, models are formulated from existing biochemical knowledge, and then model parameters are estimated using experimental data [202–204]. Parameter estimation is typically framed as a non-linear optimization problem wherein the residual (or objective function) between experimental measurements and model simulations is minimized using an optimization strategy [205]. Optimal parameter estimates are then used to predict unseen experimental data. If the validation studies fail, model construction and calibration are repeated iteratively until satisfactory results are obtained. As our biological knowledge increases, model formulation may not be as significant a challenge, but parameter estimation will likely remain difficult.

Parameter estimation is a major challenge to the development of biochemical models. Parameter estimation has been a well studied engineering problem for decades [206–209]. However, the complex dynamics of large biological systems and noisy, often incomplete experimental data sets pose a unique estimation challenge. Often optimization problems involving biological systems are non-linear and multi-modal i.e., typical models have multiple local minima or maxima [203, 205]. Non-linearity coupled with multi-modality

renders local optimization techniques such as pattern search [210], Nelder-Mead simplex methods [211], steepest descent or Levenberg-Marquardt [212] incapable of reliably obtaining globally optimal solutions as these methods often terminate at local minimum. Though deterministic global optimization techniques (for example algorithms based on branch and bound) can handle non-linearity and multi-modality [213, 214], the absence of derivative information, discontinuous objective functions, non-smooth regions or the lack of knowledge about the objective function hampers these techniques.

Meta-heuristics like Genetic Algorithms (GAs) [215], Simulated Annealing (SA) [216], Evolutionary Programming [217] and Differential Evolution (DE) [218–221] have all shown promise on non-linear multi-modal problems [222]. These techniques do not make any assumptions, nor do they require, *a priori* information about the structure of the objective function. Meta-heuristics are often very effective at finding globally optimal or near optimal solutions. For example, Mendes et al. used SA to estimate rate constants for the inhibition of HIV proteinase [223], while Modchang et al. used a GA to estimate parameters for a model of G-protein-coupled receptor (GPCR) activity [224]. Parameter estimates obtained using the GA stratified the effectiveness of two G-protein agonists, N6-cyclopentyladenosine (CPA) and 5'-N-ethylcarboxamidoadenosine (NECA). Tashkova et al. compared different meta-heuristics for parameter estimation on a dynamic model of endocytosis; DE was the most effective of the approaches tested [225]. Banga and co-workers have also successfully applied scatter-search to estimate model parameters [226–228]. Hybrid approaches, which combine meta-heuristics with local optimization techniques, have also become popular. For example, Villaverde et al. developed the enhanced scatter search (eSS) method [229], which combined scatter and local search methods, for parameter estimation in biological models [230]. However, despite these successes, a major drawback of most meta-heuristics remains the large number of function evaluations

required to explore parameter space. Performing numerous potentially expensive function evaluations is not desirable (and perhaps not feasible) for many types of biochemical models. Alternatively, Tolson and Shoemaker found, using high-dimensional watershed models, that perturbing only a subset of parameters was an effective strategy for estimating parameters in expensive models [21]. Their approach, called Dynamically Dimensioned Search (DDS), is a simple stochastic single-solution heuristic that estimates nearly optimal solutions within a specified maximum number of function (or model) evaluations. Thus, while meta-heuristics are often effective at estimating globally optimal or nearly optimal solutions, they require a large number of function evaluations to converge to a solution.

In this study, we developed Dynamic Optimization with Particle Swarms (DOPS), a novel hybrid meta-heuristic that combines the global search capability of multi-swarm particle swarm optimization with the greedy refinement of dynamically dimensioned search (DDS). The objective of DOPS is to obtain near optimal parameter estimates for large biochemical models within a relatively few function evaluations. DOPS uses multi-swarm particle swarm optimization to generate nearly optimal candidate solutions, which are then greedily updated using dynamically dimensioned search. We tested DOPS using a combination of classic optimization test functions, biochemical benchmark problems and real-world biochemical models. First, we tested the performance of DOPS on the Ackley and Rosenbrock functions, and published biochemical benchmark problems. Next, we used DOPS to estimate the parameters of a model of the human coagulation cascade. On average, DOPS outperformed other common meta-heuristics like differential evolution, simulated annealing, single-swarm particle swarm optimization, and dynamically dimensioned search on the optimization test functions, benchmark problems and the coagulation model. For example, DOPS recovered the nominal parameters for

the benchmark problems using an order of magnitude fewer function evaluations than eSS in all cases. It also produced parameter estimates for the coagulation model that predicted unseen coagulation data sets. Thus, DOPS is a promising hybrid meta-heuristic for the estimation of biochemical model parameters in relatively few function evaluations.

6.2 Results

DOPS explores parameter space using a combination of global methods.

DOPS is a novel hybrid meta-heuristic which combines a multi-swarm particle swarm method with the dynamically dimensioned search approach of Shoemaker and colleagues (Fig. 6.3).

The goal of DOPS is to estimate optimal or near optimal parameter vectors for high-dimensional biological models within a specified number of function evaluations. Toward this objective, DOPS begins by using a multi-swarm particle swarm search and then dynamically switches, using an adaptive switching criteria, to the DDS approach. The particle swarm search uses multiple sub-swarms wherein the update to each particle (corresponding to a parameter vector estimate) is influenced by the best particle amongst the sub-swarm, and the current globally best particle. Particle updates occur within sub-swarms for a certain number of function evaluations, after which the sub-swarms are reorganized. This sub-swarm mixing is similar to the regrouping strategy described by Zhao et al. [231]. DOPS switches out of the particle swarm phase based upon an adaptive switching criteria that is a function of the rate of error convergence. If the error represented by the best particle does decrease for a threshold number of function evaluations, DOPS switches automatically to the DDS search phase. The DDS search is initialized with the globally best particle from the particle swarm phase, thereafter, the particle is greedily updated by perturbing a subset of dimensions for the remaining number of function evaluations. The identity of the parameters perturbed is chosen randomly, with fewer parameters perturbed the higher the number of function evaluations.

6.2.1 DOPS minimized benchmark problems using fewer function evaluations.

On average, DOPS performed similarly or outperformed the four other meta-heuristics for the Ackley and Rastrigin test functions (Fig. 6.4). The Ackley and Rastrigin functions both have multiple local extrema and attain a global minimum value of zero.

In each case, we fixed the maximum number of function evaluations at $\mathcal{N} = 4000$ and ran $\mathcal{T} = 25$ independent experiments with different initial parameter vectors. DOPS found optimal or near optimal solutions for both the 10-dimensional Ackley (Fig. 6.4A) and Rastrigin (Fig. 6.4B) functions within the budget of function evaluations. In each of the 10-dimensional cases, other meta-heurtistics such as DDS and DE also performed well. However, DOPS consistently outperformed all other approaches tested. This performance difference was more pronounced as the dimension of the search problem increased; for a 300-dimensional Rastrigin function, DOPS was the only approach to find an optimal or near optimal solution within the function evaluation budget (Fig. 6.4B). DOPS also had a lower variability in the best value obtained (Fig. A.8), faster convergence (Fig. A.5, Fig. A.6) and less stochasticity (Fig. A.7) across multiple runs when compared to other metaheuristics. We also checked for statistical significance by doing a pairwise 2-sample t-test at 0.5% significance level. In each case we found that the p-value was close to zero disproving the null hypothesis and confirming the statistical significance of our results. Taken together, DOPS performed at least as well as other meta-heuristic approaches on small dimensional test problems, but seemed especially suited to large dimensional search spaces. Next, we tested DOPS on benchmark biochemical models of varying complexity.

Villaverde and co-workers published a set of benchmark biochemical problems to evaluate parameter estimation methods [230]. They ranked the example problems by computational cost from most to least expensive. We evaluated the performance of DOPS on problems from the least and most expensive categories. The least expensive problem (henceforth referred to as CHO model) was a metabolic model of Chinese Hamster Ovary (CHO) with 35 metabolites, 32 reactions and 117 parameters [232]. The biochemical reactions were modeled using modular rate laws and generalized Michaelis-Menten kinetics. On the other hand, the expensive problem was a genome scale kinetic model of *Saccharomyces cerevisiae* with 261 reactions, 262 variables and 1759 parameters [233] (henceforth referred to as yeast model). In both cases, synthetic time series data generated with known parameter values, was used to estimate model parameters. For yeast model, the time series data consisted of 44 observables, and for problem B4 the data corresponded to 13 different metabolite measurement sets. We fixed the number of function evaluations at $N = 4000$, and trained both models against the synthetic experimental data. DOPS produced good fits to the synthetic data (Fig. A.1 and Fig. A.2), and recapitulated the nominal parameter values using only $N \leq 4000$ function evaluations (Fig. A.4). On the other hand, enhanced scatter search (eSS) with a local optimizer took on order 10^5 function evaluations for the same problems. Thus, DOPS estimated the parameters in benchmark biochemical models, and recovered the original parameters from synthetic data, using fewer function evaluations. Next, we compared the performance of DOPS with the four other meta-heuristics for a model of the human coagulation cascade.

6.2.2 DOPS estimated the parameters of a human coagulation model.

Coagulation is an archetype biochemical network that is highly interconnected, containing both negative and positive feedback (Fig. 6.5). The biochemistry of coagulation, though complex, has been well studied [234–240], and reliable experimental protocols have been developed to interrogate the system [241–244]. Coagulation is mediated by a family proteases in the circulation, called factors and a key group of blood cells, called platelets. The central process in coagulation is the conversion of prothrombin (fII), an inactive coagulation factor, to the master protease thrombin (FIIa). Thrombin generation involves three phases, initiation, amplification and termination. Initiation requires a trigger event, for example a vessel injury which exposes tissue factor (TF), which leads to the activation of factor VII (FVIIa) and the formation of the TF/FVIIa complex. Two converging pathways, the extrinsic and intrinsic cascades, then process and amplify this initial coagulation signal. There are several control points in the cascade that inhibit thrombin formation, and eventually terminate thrombin generation. Tissue Factor Pathway Inhibitor (TFPI) inhibits upstream activation events, while antithrombin III (ATIII) neutralizes several of the proteases generated during coagulation, including thrombin. Thrombin itself also inadvertently plays a role in its own inhibition; thrombin, through interaction with thrombomodulin, protein C and endothelial cell protein C receptor (EPCR), converts protein C to activated protein C (APC) which attenuates the coagulation response by proteolytic cleavage of amplification complexes. Termination occurs after either prothrombin is consumed, or thrombin formation is neutralized by inhibitors such as APC or ATIII. Thus, the human coagulation cascade is an ideal test case; coagulation is challenging because it contains both fast and slow dynamics, but also accessible because of the availability of comprehensive data sets for model identification and validation. In this study, we used the coagulation model of Luan et al. [244], which is a coupled system of non-linear

ordinary differential equations where biochemical interactions were modeled using mass action kinetics. The Luan model contained 148 parameters and 92 species and has been validated using 21 published experimental datasets.

DOPS estimated the parameters of a human coagulation model for TF/VIIa initiated coagulation without anticoagulants (Fig. 6.7). The objective function was an unweighted linear combination of two error functions, representing coagulation initiated with different concentrations of TF/FVIIa (5pM, 5nM) [241]. We restricted the number of function evaluations to $\mathcal{N}$ = 4000 for each algorithm we tested, and performed $\mathcal{T}$ = 25 trials of each experiment to collect average performance data (Table 6.1). DOPS converged faster and had a lower final error compared to the other algorithms (Fig. 6.6). Within the first 25% of function evaluations, DOPS produced a rapid drop in error followed by a slower but steady decline. Approximately between 500-1000 function evaluations DOPS switched to the dynamically dimensioned search phase, however this transition varied from trial to trial since the switch was based upon the local convergence rate. On average, DOPS minimized the coagulation model error to a greater extent than the other meta-heuristics. However, it was unclear if the parameters estimated by DOPS had predictive power on unseen data. To address this question, we used the final parameters estimated by DOPS to simulate data that was not used for training (coagulation initiated with 500pM, 50pM, and 10pM TF/VIIa). The optimal or near optimal parameters obtained by DOPS predicted unseen coagulation datasets (Fig. 6.8). The normalized standard error for the coagulation predictions was consistent with the training error, with the exception of the 50pM TF/VIIa case which was a factor 2.75 worse (Table 6.2). Coagulation initiation with 50pM TF/FVIIa was the farthest away from the training conditions. However, while the 50pM TF/FVIIa case was the worst of those tested, it still had an average error rate of only 11% (which is near the margin of error for the experimental measurements). Taken

together, DOPS estimated parameter sets with predictive power on unseen coagulation data using fewer function iterations than other meta-heuristics. Next, we explored how the number of sub-swarms and the switch to DDS influenced the performance of the approach.

Phase switching was critical to DOPS performance.

A differentiating feature of DOPS is the switch to dynamically dimensioned search following stagnation of the initial particle swarm phase. We quantified the influence of the number of sub-swarms and the switch to DDS on error convergence by comparing DOPS with and without DDS for different numbers of sub-swarms (Fig. 6.9). We considered multi swarm particle swarm optimization with and without the DDS phase for $\mathcal{N} = 4000$ function evaluations and $\mathcal{T} = 25$ trials on the coagulation model. We used one, two, four, five and eight sub-swarms, with a total of 40 particles divided evenly amongst the swarms. Hence, we did not consider swarm numbers of three and seven. All other algorithm parameters remained the same for all cases. Generally, the higher sub-swarm numbers converged in fewer function evaluations, where the optimum particle partitioning was in the neighborhood of five sub-swarms. However, the difference in convergence rate was qualitatively similar for four, five and eight sub-swarms, suggesting there was an optimal number of particles per swarm beyond which there was no significant advantage. The multi-swarm particle swarm optimization stagnated after 25% of the available function evaluations irrespective of the number of sub-swarms. However, DOPS (with five sub-swarms) switched to DDS after detecting the stagnation. The DDS phase refined the globally best particle to produce significantly lower error on average when compared to multi-swarm particle swarm optimization alone. Thus, the automated switching strat-

egy was critical to the overall performance of DOPS.

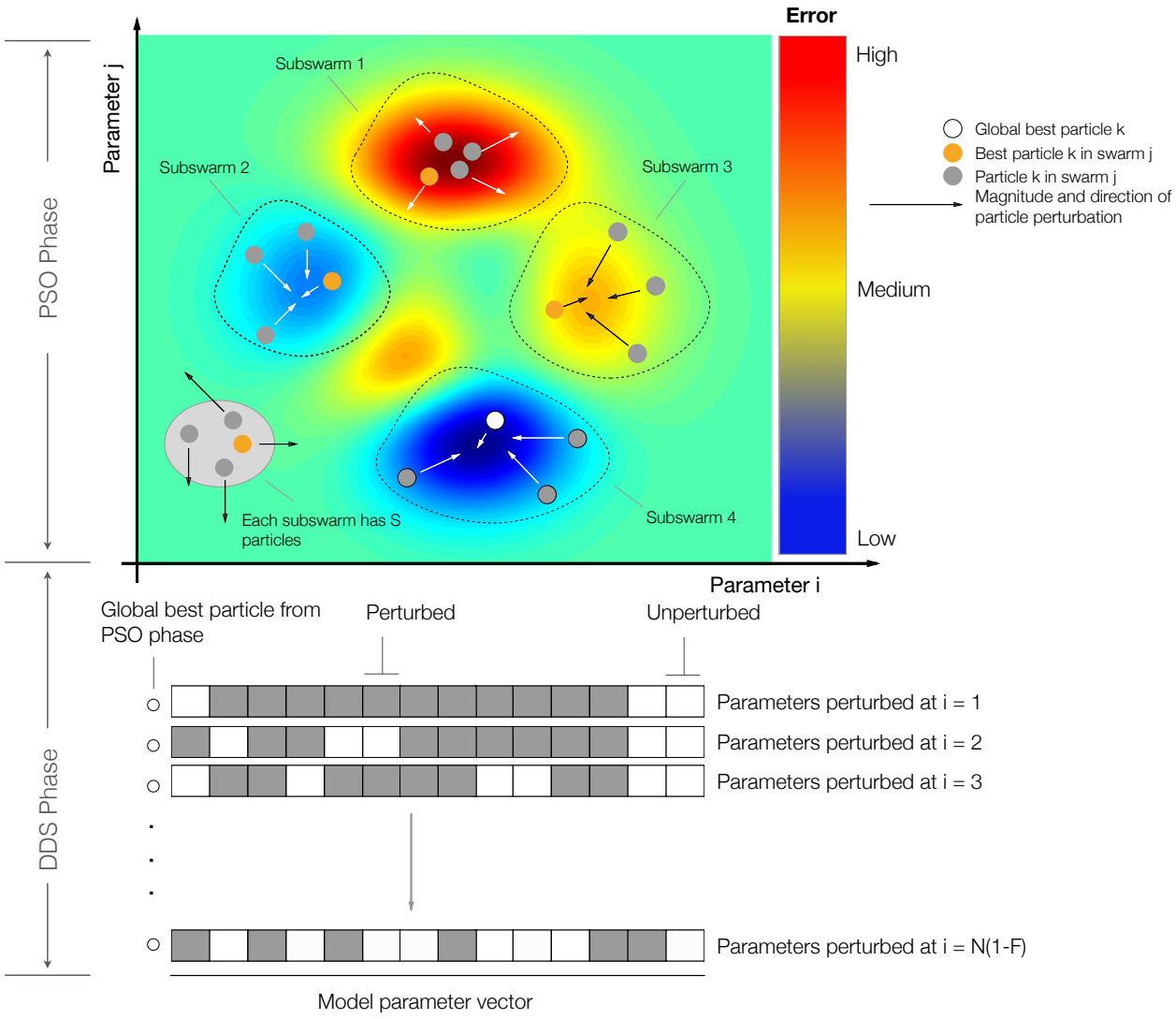

Figure 6.1: Schematic of the dynamic optimization with particle swarms (DOPS) approach. **A**: Each particle represents an N dimensional parameter vector. Particles are given randomly generated initial solutions and grouped into different sub-swarms. Within each swarm the magnitude and direction of the movement a particle is influenced by the position of the best particle and also by its own experience. After every **g** number of function evaluations the particles are mixed and randomly assigned to different swarms. When the error due to the global best particle (best particle amongst all the sub-swarms) does not drop over a certain number of function evaluations, the swarm search is stopped and the search switches to a Dynamically Dimensioned Search with global best particle as the initial solution vector or candidate vector. **B**: The candidate vector performs a greedy global search for the remaining number of function evaluations. The search neighborhood is dynamically adjusted by varying the number of dimensions that are perturbed (in black) in each evaluation step. The probability that a dimension is perturbed decreases as the number of function evaluations increase.

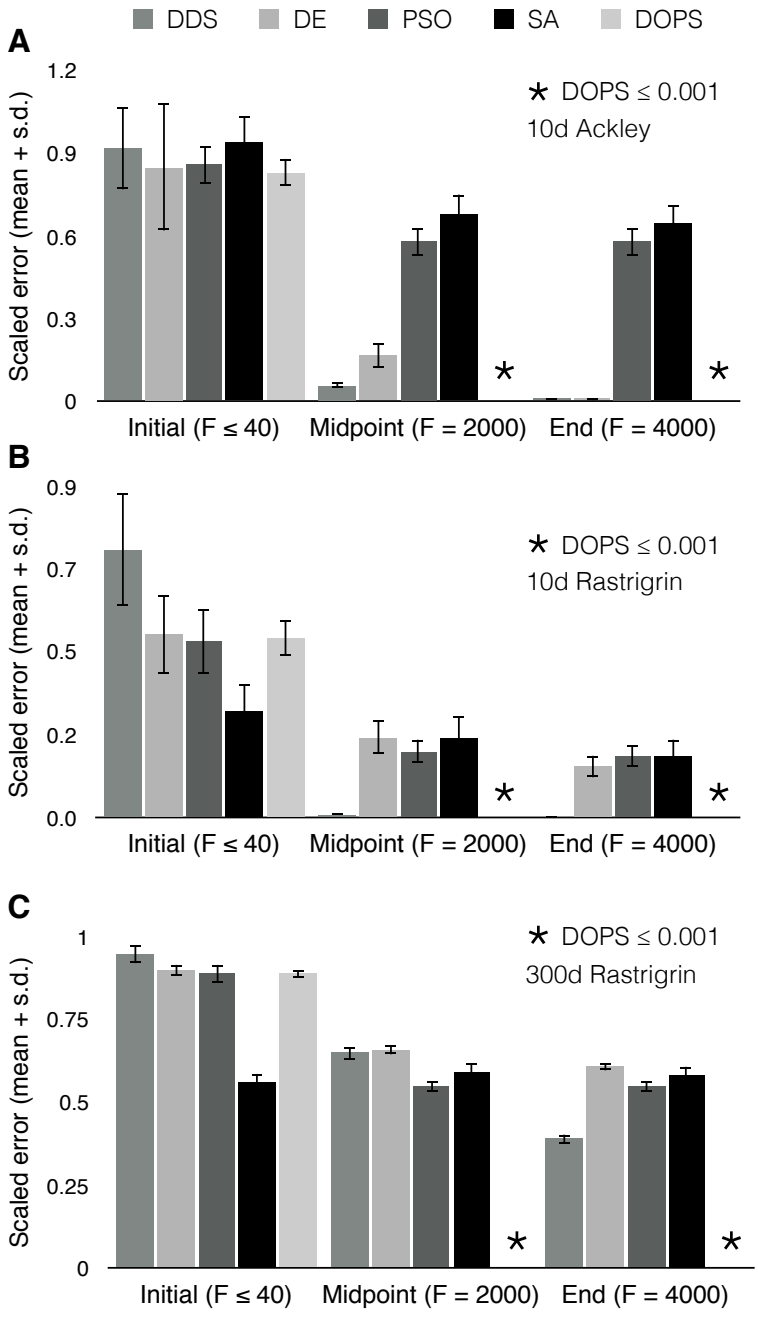

Figure 6.2: Performance of DOPS and other meta-heuristics for the Ackley and Rastrigin functions. **A**: Mean scaled error versus the number of function evaluations for the 10-dimensional Ackley function. DOPS, DDS and DE find optimal or near optimal solutions within the specified number of function evaluations. **B**: Mean scaled error versus the number of function evaluations for the 10-dimensional Rastrigin function. DOPS and DDS find optimal or near optimal solutions within the specified number of function evaluations. **C**: Mean scaled error versus the number of function evaluations for the 300-dimensional Rastrigin function. DOPS is the only algorithm that finds an optimal or near optimal solution within the specified number of function evaluations. In all cases, the maximum number of function evaluations was $\mathcal{N} = 4000$. Mean and standard deviation were calculated over $\mathcal{T} = 25$ trials.

6.3 Discussion

In this study, we developed dynamic optimization with particle swarms (DOPS), a novel meta-heuristic for parameter estimation in models of biological systems. DOPS combined multi-swarm particle swarm optimization, a global search approach, with the greedy strategy of dynamically dimensioned search to estimate optimal or nearly optimal solutions in a fixed number of function evaluations. We tested the performance of DOPS and four widely used meta-heuristics on the Ackley and Rastrigin test functions, a set of biochemical benchmark problems and a model of the human coagulation cascade. As the number of parameters increased, DOPS outperformed the other meta-heuristics, generating optimal or nearly optimal solutions using significantly fewer function evaluations compared with the other methods. We tested the solutions generated by DOPS by comparing the estimated and true parameters in the benchmark studies, and by using the coagulation model to predict unseen experimental data. For both benchmark problems, DOPS retrieved the true parameters in significantly fewer function evaluations than other meta-heuristics. For the coagulation model, we used experimental coagulation measurements under two different conditions to estimate optimal or nearly optimal parameters. These parameters were then used to predict unseen coagulation data; the coagulation model parameters estimated by DOPS predicted the correct thrombin dynamics following TF/FVIIa induced coagulation without anticoagulants. Lastly, we showed the average performance of DOPS improved when combined with dynamically dimensioned search phase, compared to an identical multi-swarm approach alone. Taken together, DOPS is a promising meta-heuristic for the estimation of parameters in large biochemical models.

Meta-heuristics can be effective techniques to estimate optimal or nearly optimal so-

lutions for complex, multi-modal functions. However, meta-heuristics typically require a large number of function evaluations to converge to a solution. DOPS is a combination of particle swarm optimization, which is a global search method, and dynamically dimensioned search, which is a greedy evolutionary technique. Particle swarm optimization uses collective information shared amongst swarms of computational particles to search for global extrema. Several particle swarm variants have been proposed to improve the search ability and rate of convergence. These variations involve different neighborhood structures, multi-swarms or adaptive parameters. Multi-swarm particle swarm optimization with small particle neighborhoods has been shown to be better in searching on complex multi-modal solutions [231]. Multi-swarm methods generate diverse solutions, and avoid rapid convergence to local optima. However, at least for the coagulation problem used in this study, multi-swarm methods stagnated after approximately 25% of the available function evaluations; only the introduction of dynamically dimensioned search improved the rate of error convergence. Dynamically dimensioned search, which greedily perturbs only a subset of parameter dimensions in high dimensional parameter spaces, refined the globally best particle and produced significantly lower error on average when compared to multi-swarm particle swarm optimization alone. However, dynamically dimensioned search, starting from a initial random parameter guess, was not as effective on average as DOPS. The initial solutions generated by the multi swarm search had a higher propensity to produce good parameter estimates when refined by dynamically dimensioned search. Thus, our hybrid combination of two meta-heuristics produced better results than either constituent approach, and better results than other meta-heuristic approaches on average. This was true of not only the convergence rate on the coagulation problem, but also the biochemical benchmark problems; DOPS required two-orders of magnitude fewer function evaluations compared with enhanced Scatter Search (eSS) to estimate the biochemical benchmark model parameters. Taken together, the combina-

tion of particle swarm optimization and dynamically dimensioned search produced better results than either of these constituent approaches alone, and required fewer function evaluations compared with other common meta-heuristics.

DOPS performed well on many different systems with no pre-optimization of algorithm parameters, however there are many research questions that should be pursued further. DOPS comfortably outperformed existing, widely used meta-heuristics for high dimensional global optimization functions, biochemical benchmark models and a model of the human coagulation system. However, it is possible that highly optimized versions of common meta-heuristics could surpass DOPS; we should compare the performance of DOPS with optimized versions of the common meta-heuristics on both test and real-world problems to determine if a performance advantage exists in practice. Next, DOPS has a hybrid architecture, thus the particle swarm phase could be combined with other search strategies such as local derivative based approaches to improve convergence rates. We could also consider multiple phases beyond particle swarm and dynamically dimensioned search, for example switching to a gradient based search following the dynamically dimensioned search phase. Next, the current implementation of the algorithm switches only once from the swarm phase to the dynamically dimensioned phase. An obvious modification would be to allow multiple transitions between the swarm and dynamically dimensioned phases. Lastly, we should update DOPS to treat multi-objective problems. The identification of large biochemical models sometimes requires training using qualitative, conflicting or even contradictory data sets. One strategy to address this challenge is to estimate experimentally constrained model ensembles using multi-objective optimization. Previously, we developed Pareto Optimal Ensemble Techniques (POETs) which integrates simulated annealing with Pareto optimality to identify models near the optimal tradeoff surface between competing training objectives [245]. Since

DOPS consistently outperformed simulated annealing on both test and real-world problems, we expect a multi-objective form of DOPS would more quickly estimate solutions which lie along high dimensional trade-off surfaces.

6.4 Materials and Methods

Optimization problem formulation.

Model parameters were estimated by minimizing the difference between model simulations and $\mathcal{E}$ experimental measurements. Simulation error is quantified by an objective function $K(\mathbf{p})$ (typically the Euclidean norm of the difference between simulations and measurements) subject to problem and parameter constraints:

$$\begin{aligned} \min_{\mathbf{p}} K(\mathbf{p}) \quad &= \sum_{i=1}^{\mathcal{E}} (g_i(t_i, \mathbf{x}, \mathbf{p}, \mathbf{u}) - y_i)^2 \\ \text{subject to} \quad &\dot{\mathbf{x}} = \mathbf{f}(t, \mathbf{x}(t, \mathbf{p}), \mathbf{u}(t), \mathbf{p}) \\ &\mathbf{x}(t_0) = \mathbf{x}_0 \\ &\mathbf{c}(t, \mathbf{x}, \mathbf{p}, \mathbf{u}) \geqslant \mathbf{0} \\ &\mathbf{p}^L \leqslant \mathbf{p} \leqslant \mathbf{p}^U \end{aligned} \tag{6.1}$$

The term $K(\mathbf{p})$ denotes the objective function, t denotes time, $g_i(t_i, \mathbf{x}, \mathbf{p}, \mathbf{u})$ is the model output for experiment i, $\mathbf{x}(t, \mathbf{p})$ is the state variable vector with an initial state $\mathbf{x}_0$, $\mathbf{u}(t)$ is a model input vector, $\mathbf{f}(t, \mathbf{x}(t, \mathbf{p}), \mathbf{u}(t), \mathbf{p})$ is the system of model equations (e.g., differential equations or algebraic constraints) and $\mathbf{p}$ denotes the model parameter vector. The parameter search (or model simulations) can be subject to $\mathbf{c}(t, \mathbf{x}, \mathbf{p}, \mathbf{u})$ linear or non-linear constraints, and parameter bound constraints where $\mathbf{p}^L$ and $\mathbf{p}^U$ denote the lower and upper parameter bounds, respectively. Optimal model parameters are then given by:

$$\mathbf{p}^* = \arg\min_{\mathbf{p}} K(\mathbf{p}) \tag{6.2}$$

6.4.1 Dynamic optimization with particle swarms (DOPS).

DOPS is a novel meta-heuristic which combines multi-swarm particle swarm optimization with the dynamically dimensioned search (Fig. 6.3) and (Algo. 1). Particle swarm optimization is an evolutionary algorithm that uses a population of particles (solutions) to find an optimal solution. Each particle is updated based on its on experience (particle best) and the experience of all other particles within the swarm (global best) [246][247][248]. Multi-swarm particle swarm optimization is a variant of particle swarm optimization that uses multiple swarms instead of a single swarm. DOPS uses a novel update rule of particles within each swarm. This rule uses only particle positions but not particle velocities in the update rule. This differs from existing PSO methods that rely both on particle position and velocity for update. The update rule further described in detail in the methods. The goal of DOPS is to estimate optimal or near optimal parameter vectors for high-dimensional biological models within a specified number of function evaluations. Toward this objective, DOPS begins by using a particle swarm search and then dynamically switches, using an adaptive switching criteria, to a DDS search phase. Dynamically Dimensioned Search (DDS) is a single solution based search algorithm. It was developed by Tolson and Shoemaker to obtain good solutions to high-dimensional search problems within a fixed number of function evaluations. DDS starts as a global search algorithm by perturbing all the dimensions. Later the number of dimensions that are perturbed is decreased with a certain probability. The probability that a certain dimension is perturbed reduces (a minimum of one dimension is always perturbed) as the iterations increase. This causes the algorithm to behave as a local search algorithm as the number of iterations increase. The perturbation magnitude of each dimension is from normal distribution with zero mean. The standard deviation that was used in the original DDS paper and the current study is 0.2. DDS performs a greedy search where the solution

is updated only if it is better than the previous solution. The combination of perturbing a subset of dimensions along with greedy search indirectly relies on model sensitivity to a specific parameter combination. For the sake of brevity we are not providing the entire pseudo code of DDS here. The reader is requested to refer to Figure 1 of the original paper by Tolson and Shoemaker [21].

input : A randomized swarm of particles of size $NP \times K$ and fixed number of function evaluations N
output: Optimized parameter vector of size $1 \times K$

Initialize the particles randomly and assign particles randomly to k sub-swarms;
while $j \leq N$ **do**
  if *mod(j,G)=0* **then**
    Reassign particles to k sub-swarms;
  end
  for $i \leftarrow 1$ **to** NS **do**
    Update particles within sub-swarms according to equation 3;
  end
  Find best particle $\mathcal{G}$ *amongst all sub-swarms*;
  if $besterror(j) \geq 0.99 * besterror(j+1)$ **then**
    $failurecounter \leftarrow failurecounter + 1$;
  else
    $failurecounter \leftarrow 0$;
  end
  if $failurecounter \geq threshold$ **then**
    $\mathcal{G} \leftarrow DDS(\mathcal{G}, N - j)$;
    return $\mathcal{G}$
  else
    $j \leftarrow j + 1$;
  end
  return $\mathcal{G}$
end

Algorithm 1: Pseudo code for the dynamic optimization with particle swarms (DOPS) method.

Phase 1: Particle swarm phase.

The particle swarm phase of DOPS begins by randomly initializing a swarm of $\mathcal{K}$-dimensional particles (represented as z_i), wherein each particle corresponded to a $\mathcal{K}$-

dimensional parameter vector. After initialization, particles were randomly partitioned into k equally sized sub-swarms $\mathcal{S}_1, \ldots, \mathcal{S}_k$. Particles within each sub-swarm $\mathcal{S}_k$ were updated according to the rule:

$$z_{i,j} = \theta_{1,j-1} z_{i,j-1} + \theta_2 r_1 \left(\mathcal{L}_i - z_{i,j-1} \right) + \theta_3 r_2 \left(\mathcal{G}_k - z_{i,j-1} \right) \tag{6.3}$$

where $(\theta_1, \theta_2, \theta_3)$ were adjustable parameters, $\mathcal{L}_i$ denotes the best solution found by particle i within sub-swarm $\mathcal{S}_k$ for function evaluation $1 \rightarrow j-1$, and $\mathcal{G}_k$ denotes the best solution found over all particles within sub-swarm $\mathcal{S}_k$. The quantities r_1 and r_2 denote uniform random vectors with the same dimension as the number of unknown model parameters ($\mathcal{K} \times 1$). Equation (6.3) is similar to the general particle swarm update rule, however, it does not contain velocity terms. In DOPS, the parameter $\theta_{1,j-1}$ is similar to the inertia weight parameter for the velocity term described by Shi and Eberhart [249]; Shi and Eberhart proposed a linearly decreasing inertia weight to improve convergence properties of particle swarm optimization. Our implementation of $\theta_{1,j-1}$ is inspired by this and the decreasing perturbation probability proposed by Tolson and Shoemaker [21]. It is an analogous equivalent to inertia weight on velocity. However $\theta_{1,j-1}$ places inertia on the position rather than velocity and uses the same rule described by Shi and Eberhart to adaptively change with the number of function evaluations:

$$\theta_{1,j} = \frac{(\mathcal{N} - j) * (w_{max} - w_{min}))}{(\mathcal{N} - 1)} + w_{min} \tag{6.4}$$

where $\mathcal{N}$ represents the total number of function evaluations, w_{max} and w_{min} are the maximum and minimum inertia weights, respectively. While updating the particles, parameter bounds were enforced using reflection boundary conditions (Algo. 2).

```
if z_{i,j}^{old} < z_i^{min} then
    z_{i,j}^{new} = z_{i,j}^{old} + (z_i^{min} − z_{i,j}^{old}) if z_{i,j}^{new} > z_i^{max} then
        z_{i,j}^{new} = z_i^{max}
    end
end
if z_{i,j}^{old} > z_i^{max} then
    z_{i,j}^{new} = z_{i,j}^{old} + (z_{i,j}^{old} − z_i^{max}) if z_{i,j}^{new} < z_i^{min} then
        z_{i,j}^{new} = z_i^{min}
    end
end
```

Algorithm 2: Pseudo code for the reflective boundary conditions used by the dynamic optimization with particle swarms (DOPS) method.

After every $\mathcal{M}$ function evaluations, particles were randomly redistributed to a new sub-swarm, and updated according to Eqn. (6.3). This process continued for a maximum of $\mathcal{F} * \mathcal{N}$ functions evaluations, where $\mathcal{F}$ is the fraction of evaluations in the particle swarm phase of DOPS. However, if the simulation error stagnated e.g., did not change by more than 1% for a specified number of evaluations, the swarm phase was terminated and DOPS switched to exploring parameter space using the DDS approach.

Phase 2: DDS phase.

```
input : Candidate vector G from swarm search and (1 − F) * N evaluations
output: Optimized parameter vector of size 1 × K
while j ≤ (1 − F) * N do
    Assign probability of perturbation to each dimension P_i according to equation 7;
    Select a subset of dimensions based on a threshold value for perturbation;
    Update candidate solution G(J) according to equation 5;
    Ensure updated solution G_new(J) is within bounds using Algorithm 2;
end
```

Algorithm 3: Pseudo code for the Dynamically Dimensioned Search (DDS) method.

At the conclusion of the swarm phase, the overall best particle, $\mathcal{G}_k$, over the k subswarms was used to initialize the DDS phase. DOPS takes at least $(1 - \mathcal{F}) * \mathcal{N}$ function evaluations during the DDS phase and then terminates the search. For the DDS phase, the best parameter estimate was updated using the rule:

$$\mathcal{G}_{new}(J) = \begin{cases} \mathcal{G}(\mathbf{J}) + \mathbf{r}_{normal}(\mathbf{J})\sigma(\mathbf{J}), & \text{if } \mathcal{G}_{new}(\mathbf{J}) < \mathcal{G}(\mathbf{J}). \\ \mathcal{G}(\mathbf{J}), & \text{otherwise.} \end{cases} \tag{6.5}$$

where $\mathbf{J}$ is a vector representing the subset of dimensions that are being perturbed, $\mathbf{r}_{normal}$ denotes a normal random vector of the same dimensions as $\mathcal{G}$, and σ denotes the perturbation amplitude:

$$\sigma = R(\mathbf{p}^U - \mathbf{p}^L) \tag{6.6}$$

where R is the scalar perturbation size parameter, $\mathbf{p}^U$ and $\mathbf{p}^L$ are $(\mathcal{K} \times 1)$ vectors that represent the maximum and minimum bounds on each dimension. The set $\mathbf{J}$ was constructed using a probability function $\mathcal{P}_i$ that represents a threshold for determining whether a specific dimension j was perturbed or not; $\mathcal{P}_i$ is monotonically decreasing function of function evaluations:

$$\mathcal{P}_i = 1 - \log\left[\frac{i}{(1 - \mathcal{F}) * \mathcal{N}}\right] \tag{6.7}$$

where i is the current iteration. After $\mathcal{P}_i$ is determined, we drew $\mathcal{P}_j$ from a uniform distribution for each dimension j. If $\mathcal{P}_j < \mathcal{P}_i$ was included in $\mathbf{J}$. Thus, the probability that a dimension j was perturbed was inversely proportional to the number of function evaluations. DDS updates are greedy; $\mathcal{G}_{new}$ becomes the new solution vector only if it is better than $\mathcal{G}$. DOPS source code is available for download under a MIT license at http://www.varnerlab.org.

Acknowledgements

This study was supported by an award from the Army Research Office (ARO #59155-LS).

Table 6.1: Table with optimization settings and results for the coagulation problem, the benchmarks and test functions using DOPS. For each problem the bounds on the parameter vector, the total number of function evaluations, the best initial objective value and the best final objective value are specified. Here *pnom* indicates the nominal or true parameter vector of the model. Nominal objective value represents the objective value using the true parameter vector or the nominal parameter vector. The CPU time is the time taken for the problem on a 2.4GHz Intel Xeon Architecture running Matlab 2014b.

	Coagulation	Yeast model	CHO model	Ackley	Rastrigin
Evaluations	4000	4000	4000	4000	4000
Lower Bound	0.001.pnom	5.pnom	5.pnom	30	5.12
Upper Bound	1000.pnom	0.2.pnom	0.2.pnom	-15	-5.12
CPU Time	10.1 hrs	38.3 hrs	6.2 min	2.8 s	2.6 s
Scaled initial error	1.0	1.0	1.0	1.0	1.0
Scaled final error	< 0.01	< 0.01	< 0.01	< 0.01	< 0.01
Scaled nominal error	0.42	0.1	< 0.01	0	0

Table 6.2: Error analysis for the human coagulation model. The coagulation model was trained on coagulation initiated with TF/FVIIa at 5 nM and the 5 pM to obtain the optimal parameters. Using these optimal parameters, coagulation dynamics were predicted for varying initiator concentrations (500 pM, 50 pM and 10 pM). Model agreement with measurements was quantified using normalized squared error. The normalized squared error is defined as $N.S.E. = (1/max(\mathbf{X})) * (\|(\mathbf{Y}, \mathbf{X})\|/sqrt(\mathrm{N}))$ where $\mathbf{X}$ is the experimental data, $\mathbf{Y}$ is the model simulation data interpolated onto the experimental time scale and N is the total number of experimental time points.

TF/FVIIa concentration	Normalized S.E.	Category
5 nM	0.0376	Training
500 pM	0.0564	Prediction
50 pM	0.1125	Prediction
10 pM	0.0823	Prediction
5 pM	0.0338	Training

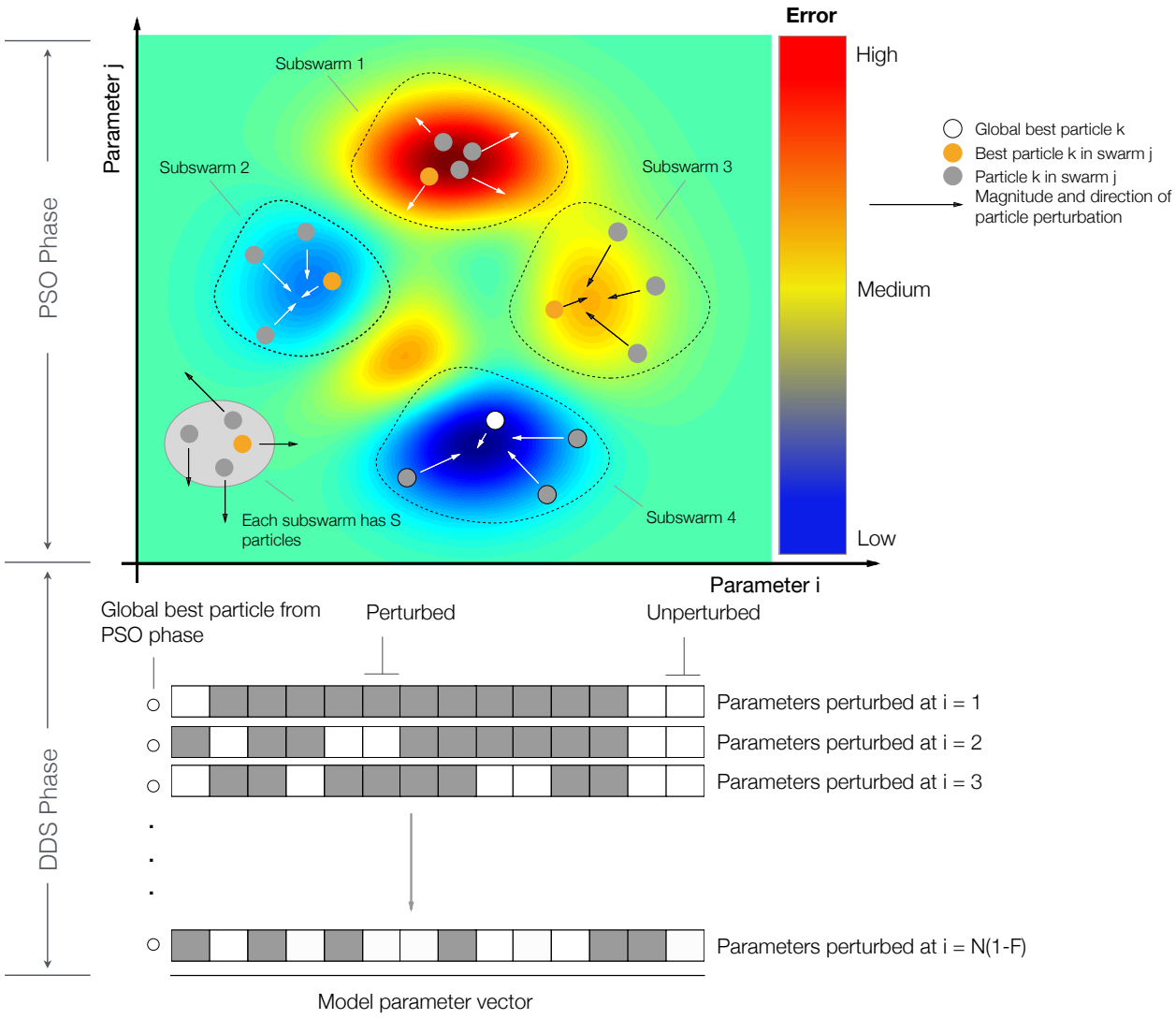

Figure 6.3: Schematic of the dynamic optimization with particle swarms (DOPS) approach. **A**: Each particle represents an N dimensional parameter vector. Particles are given randomly generated initial solutions and grouped into different sub-swarms. Within each swarm the magnitude and direction of the movement a particle is influenced by the position of the best particle and also by its own experience. After every **g** number of function evaluations the particles are mixed and randomly assigned to different swarms. When the error due to the global best particle (best particle amongst all the sub-swarms) does not drop over a certain number of function evaluations, the swarm search is stopped and the search switches to a Dynamically Dimensioned Search with global best particle as the initial solution vector or candidate vector. **B**: The candidate vector performs a greedy global search for the remaining number of function evaluations. The search neighborhood is dynamically adjusted by varying the number of dimensions that are perturbed (in black) in each evaluation step. The probability that a dimension is perturbed decreases as the number of function evaluations increase.

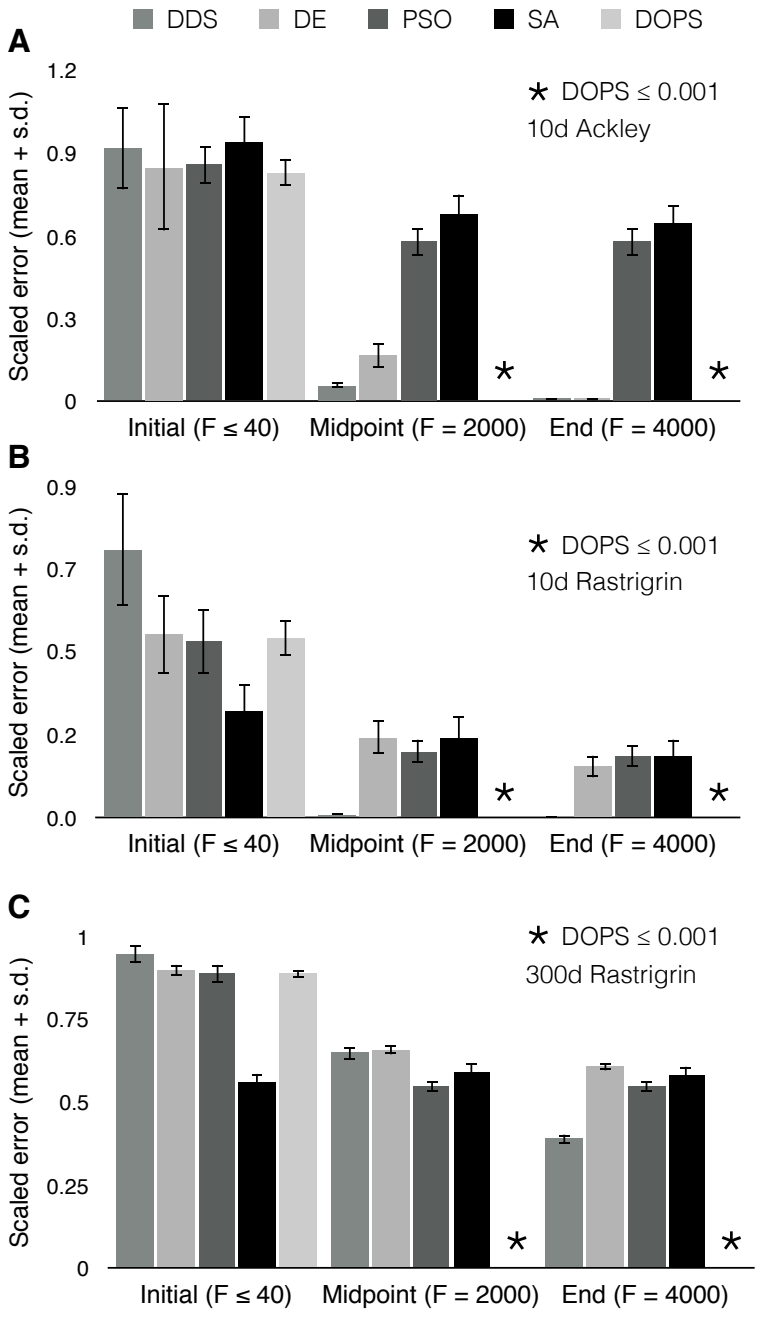

Figure 6.4: Performance of DOPS and other meta-heuristics for the Ackley and Rastrigin functions. **A**: Mean scaled error versus the number of function evaluations for the 10-dimensional Ackley function. DOPS, DDS and DE find optimal or near optimal solutions within the specified number of function evaluations. **B**: Mean scaled error versus the number of function evaluations for the 10-dimensional Rastrigin function. DOPS and DDS find optimal or near optimal solutions within the specified number of function evaluations. **C**: Mean scaled error versus the number of function evaluations for the 300-dimensional Rastrigin function. DOPS is the only algorithm that finds an optimal or near optimal solution within the specified number of function evaluations. In all cases, the maximum number of function evaluations was $\mathcal{N} = 4000$. Mean and standard deviation were calculated over $\mathcal{T} = 25$ trials.

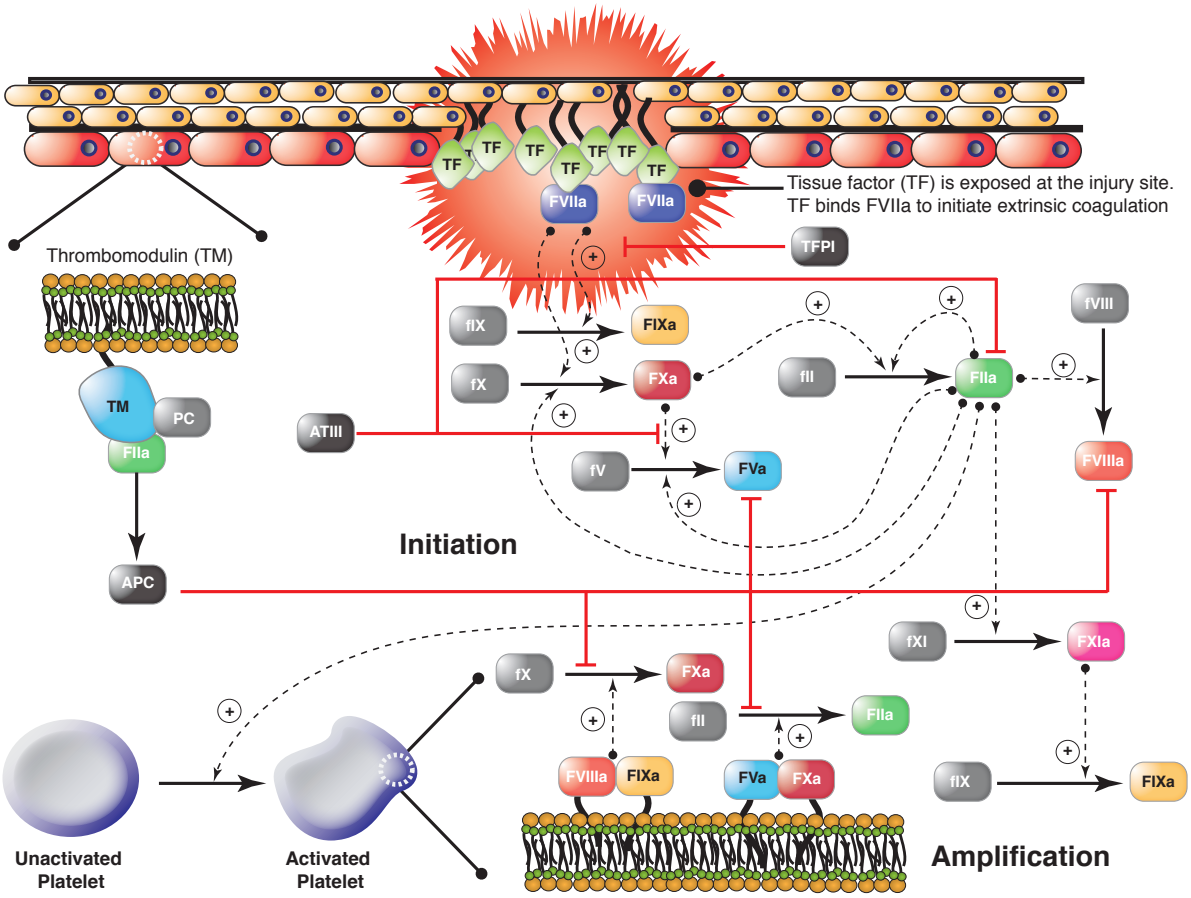

Figure 6.5: Schematic of the extrinsic and intrinsic coagulation cascade. Inactive zymogens upstream (grey) are activated by exposure to tissue factor (TF) following vessel injury. Tissue factor and activated factor VIIa (FVIIa) form a complex that activates factor X (fX) and IX (fIX). FXa activates downstream factors including factor VIII (fVIII) and fIX. Factor V (fV) is primarily activated by thrombin (FIIa). In addition, we included a secondary fV activation route involving FXa. FXa and FVa form a complex (prothrombinase) on activated platelets that converts prothrombin (fII) to FIIa. FIXa and FVIIIa can also form a complex (tenase) on activated platelets which catalyzes FXa formation. Thrombin also activates upstream coagulation factors, forming a strong positive feedback ensuring rapid activation. Tissue factor pathway inhibitor (TFPI) downregulates FXa formation and activity by sequestering free FXa and TF-FVIIa in a FXa-dependent manner. Antithrombin III (ATIII) inhibits all proteases. Thrombin inhibits itself binding the surface protein thrombomodulin (TM). The IIa-TM complex catalyzes the conversion of protein C (PC) to activated protein C (APC), which attenuates the coagulation response by the proteolytic cleavage of fV/FVa and fVIII/FVIIIa.

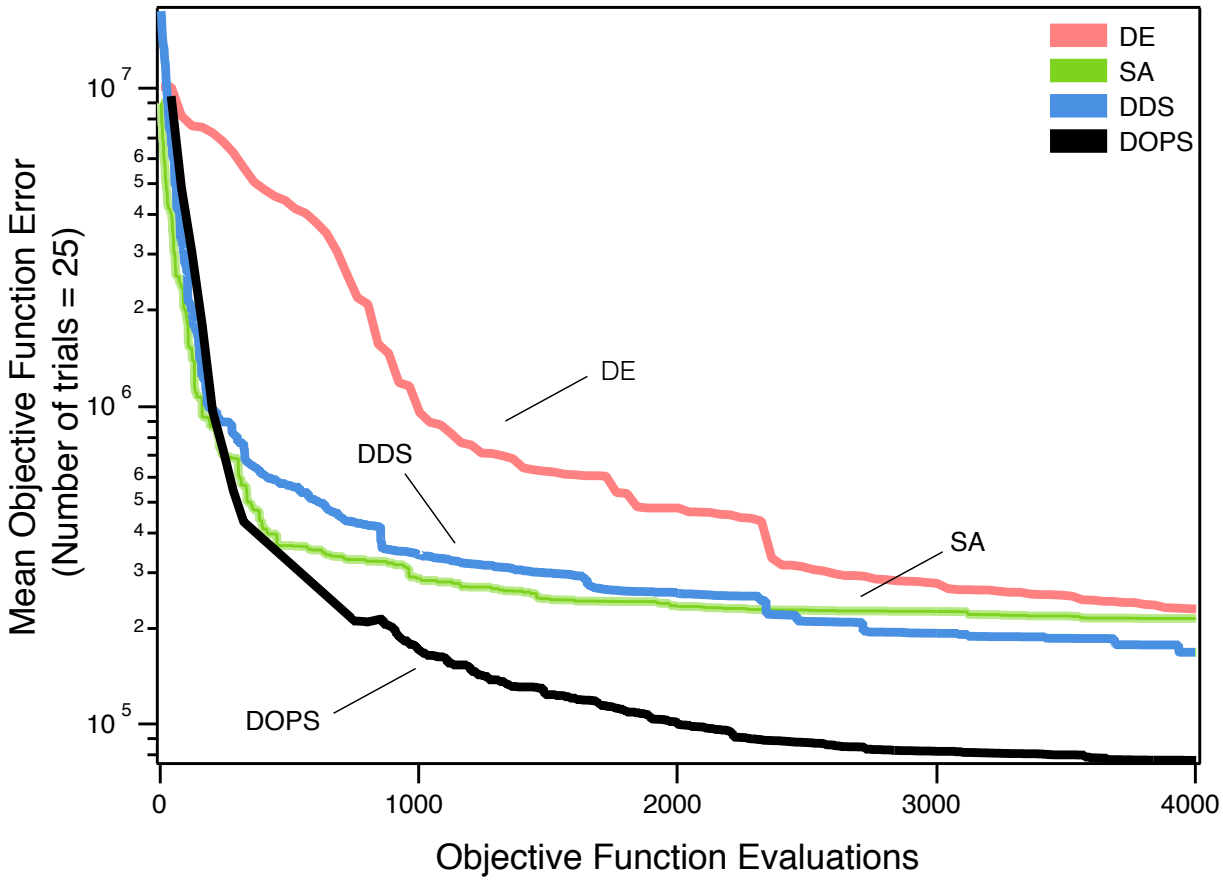

Figure 6.6: Error convergence rates of the five different algorithms on the coagulation model. The objective error is the mean over $\mathcal{T} = 25$ trials. DOPS, DDS and SA have the steepest drop in error during first 300 function evaluations. Thereafter the error drop in DDS and SA remains nearly constant whereas DOPS continues to drops further. At the end of 4000 function evaluations DOPS attains the lowest error. The next best estimate using DDS is nearly three times greater than the lowest error using DDS.

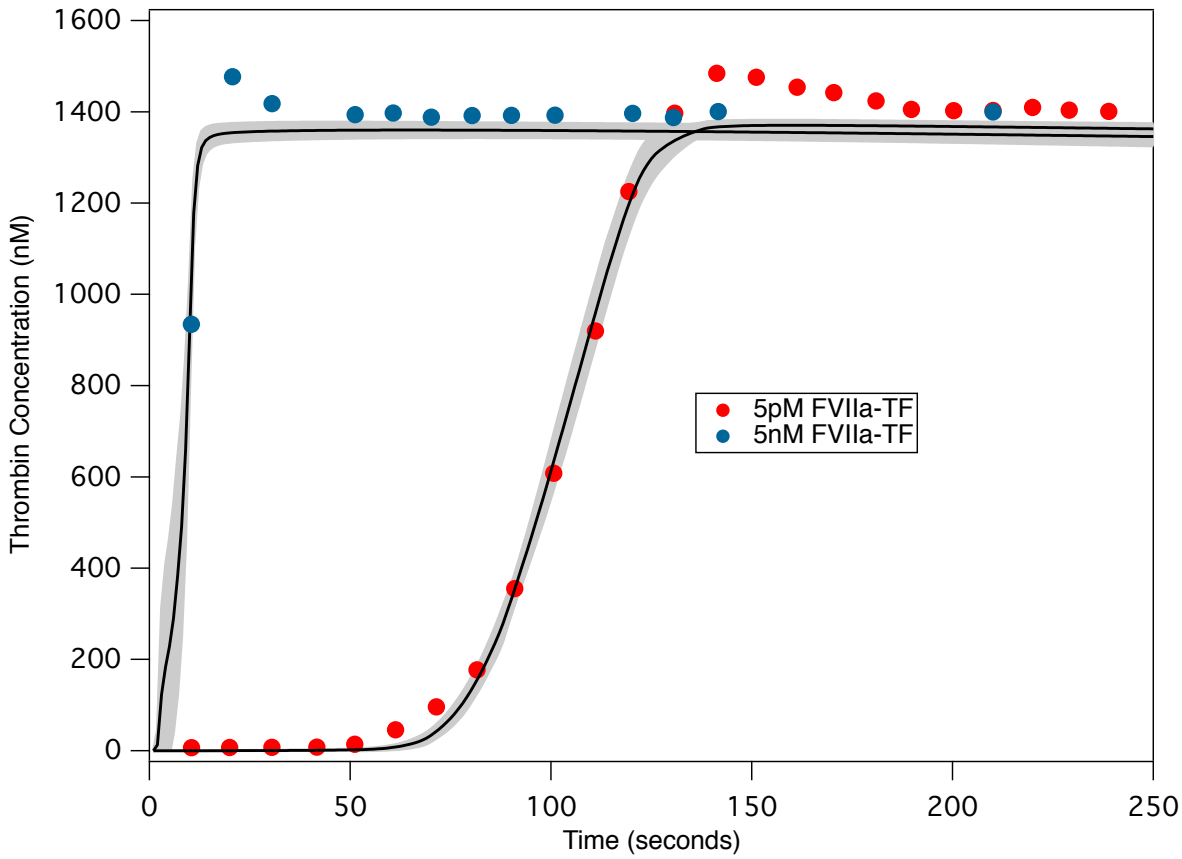

Figure 6.7: Model fits on experimental data using DOPS. The model parameters were estimated using DOPS. Solid black lines indicate the simulated mean thrombin concentration using parameter vectors from 25 trials. The grey shaded region represents the 99% confidence estimate of the mean simulated thrombin concentration. The experimental data is reproduced from the synthetic plasma assays of Mann and co-workers. Thrombin generation is initiated by adding Factor TF/VIIa (5nM and 5pM) to synthetic plasma containing 200 μmol/L of phospholipid vesicles (PCPS) and a mixture of coagulation factors (II,V,VII,VIII,IX,X and XI) at their mean plasma concentrations.

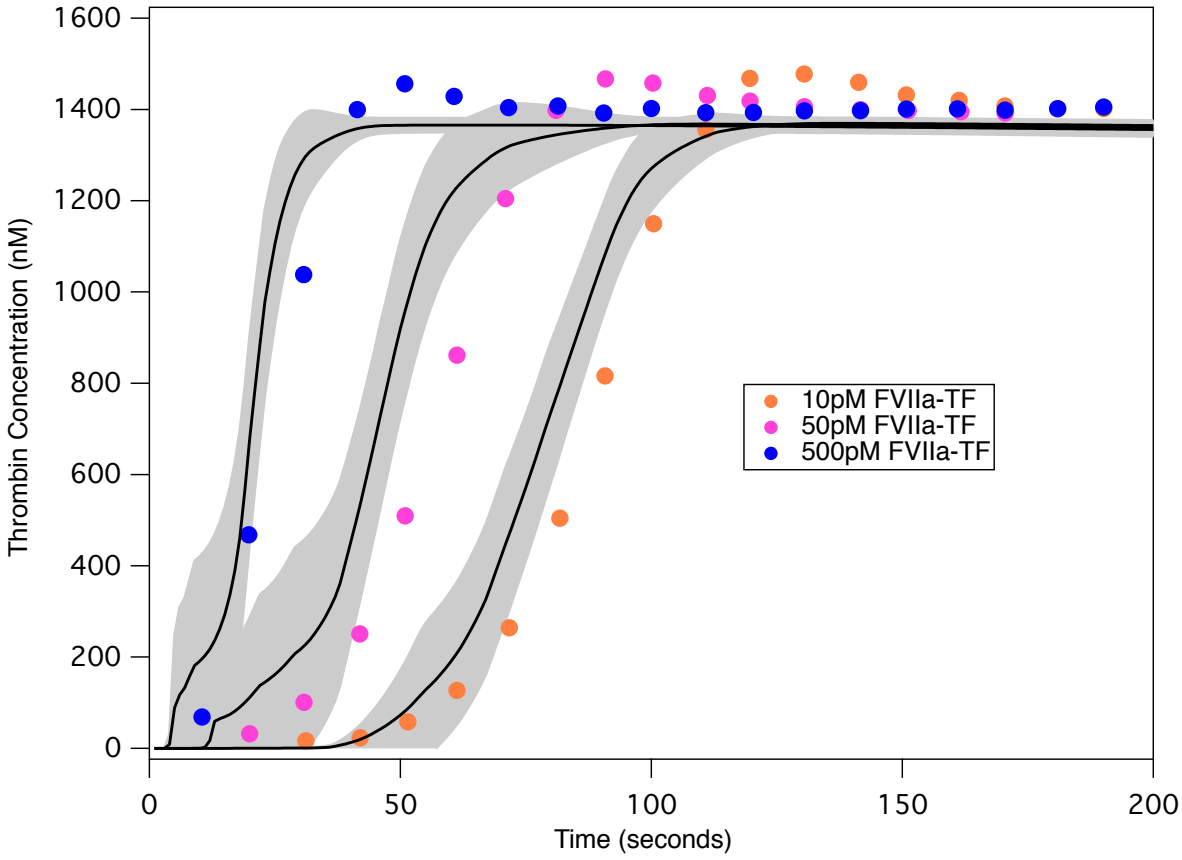

Figure 6.8: Model predictions on unseen experimental data using parameters obtained from DOPS. The parameter estimates that were obtained using DOPS were tested against data that was not used in the model training. Solid black lines indicate the simulated mean thrombin concentration using parameter vectors from $\mathcal{T} = 25$ trials. The grey shaded region represents the 99% confidence estimate of the mean simulated thrombin concentration. The experimental data is reproduced from the synthetic plasma assays of Mann and co-workers. Thrombin generation is initiated by adding Factor VIIa-TF (500pM - Blue, 50pM - Pink and 10pM - orange, respectively) to synthetic plasma containing 200 μmol/L of phospholipid vesicles (PCPS) and a mixture of coagulation factors (II,V,VII,VIII,IX,X and XI) at their mean plasma concentrations.

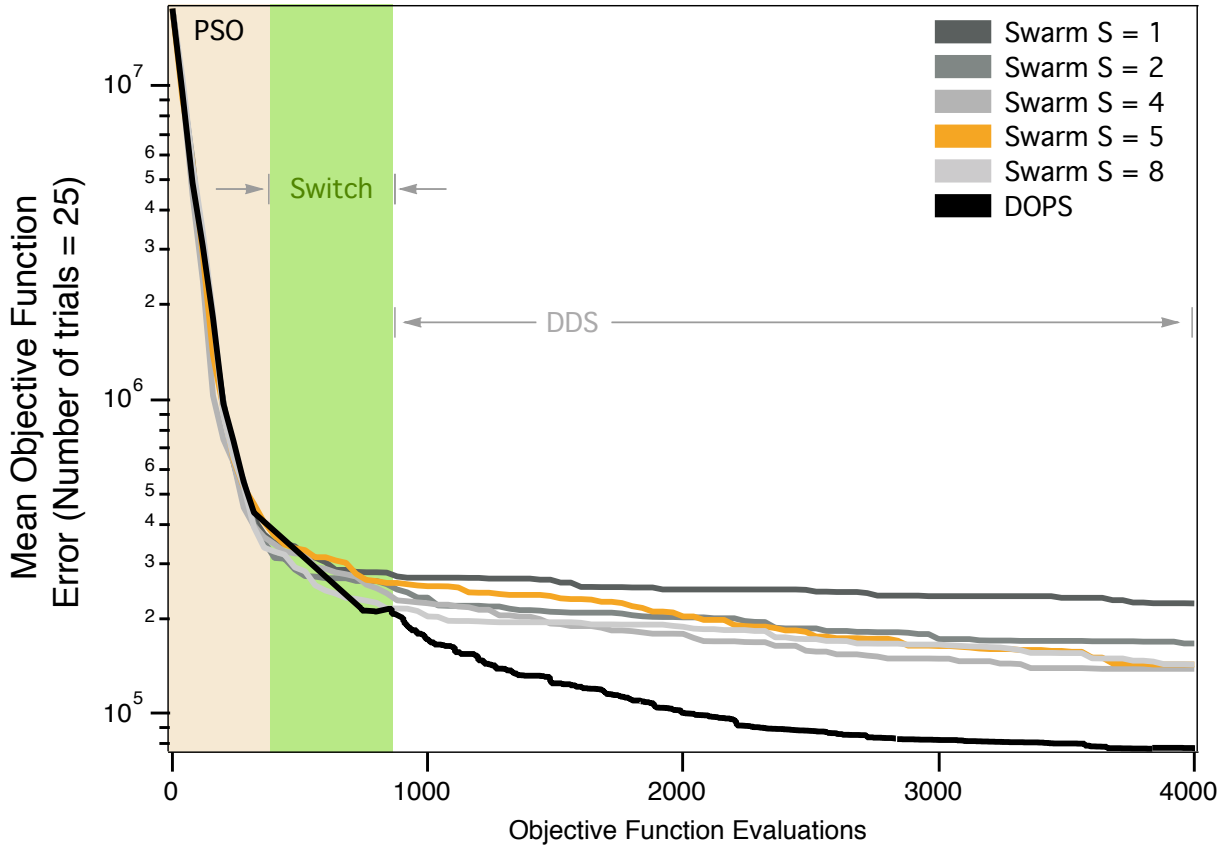

Figure 6.9: Influence of the switching strategy and sub-swarms on DOPS performance for the coagulation model. DOPS begins by using a particle swarm search and then dynamically switches (switch region), using an adaptive switching criteria, to the DDS search phase. We compared the performance of DOPS with and without DDS for different sub-swarm searches to quantify the effect of number of sub-swarms and DDS. We used one, two, four, five and eight sub-swarms, with a total of 40 particles divided evenly amongst the swarms. The results presented are the average of $\mathcal{T} = 25$ trials with $\mathcal{N} = 4000$ function evaluations each. The convergence rates with higher swarm numbers is typically higher but there is no pronounced difference amongst four, five and eight. The multi-swarm with DDS saturates while DOPS shows a rapid drop due to a switch to the DDS phase.

CHAPTER 7

SUMMARY AND FUTURE WORK

The objective of this work has been to provide a platform for building an integrated These networks are complex and often require great degree of mechanistic detail for simulation. Thus building an integrated multi-scale framework is expensive. We use an approach that reduces mechanistic detail involved by using a combination of logical rules with ODEs. In chapter 2 we illustrate this approach on a biochemical network of coagulation. We capture the dynamics of thrombin generation in vitro under different conditions. Specifically we quantify the regulatory role of protein C, a protein believed to have a crucial role in trauma induced coagulopathy. We extend this approach to model cell-free metabolic networks and captured complex allosteric patterns in chapter 3.

Chapter 4 describes perhaps the first model of complement that includes all the three pathways of initiation. Importantly we describe a model that captures dynamics of C5a, a key complement protein that interacts with proteins in coagulation and fibrinolysis under trauma. The sensitivity analysis and robustness analysis validated some of the existing current therapeutic strategies and also identified potentially important mechanisms that can be used develop novel treatment strategies. In Chapter 5 we describe a model of coagulation and fibrinolysis. The model successfully predicts ROTEM data. ROTEM data is perhaps the most important clinical metric in measuring coagulopathy. The ability to link biochemical concentrations with a clinical measure is significant and can serve as an important tool for developing new treatment strategies. We conclude these series of papers in traumatic coagulopathy with an algorithm that can train the large models of biochemical networks rapidly without problem specific parameter setting. The combination of particle swarm optimization and dynamically dimensioned search outperforms known

metaheuristics on well known benchmarks as well as a large model of coagulation. This method can utilized to obtain good starting solutions, if not optimal solutions for large models.

The future directions of this work firstly involves a building a reduced order model of coagulation that includes platelets. After building this model the next step would be to describe a whole body model using a physiology based pharmacokinetic approach using physiological parameters about organ sizes, perfusion rates. Subsequently, using the reduced order models of biochemical networks in coagulopathy dynamics of clot generation need to be captured in presence of an injury. The qualitative and quantitative predictions need to be experimentally validated using the clinical data from our collaborators.

The development and applications of whole body models is an exciting approach that has been barely exploited. It presents an immense opportunity to investigate complex biochemical and physiological phenomena. Successful application in this context will certainly provide the researchers and clinicians with novel perspectives about trauma induced coagulopathy.

APPENDIX A

CHAPTER 1 OF APPENDIX

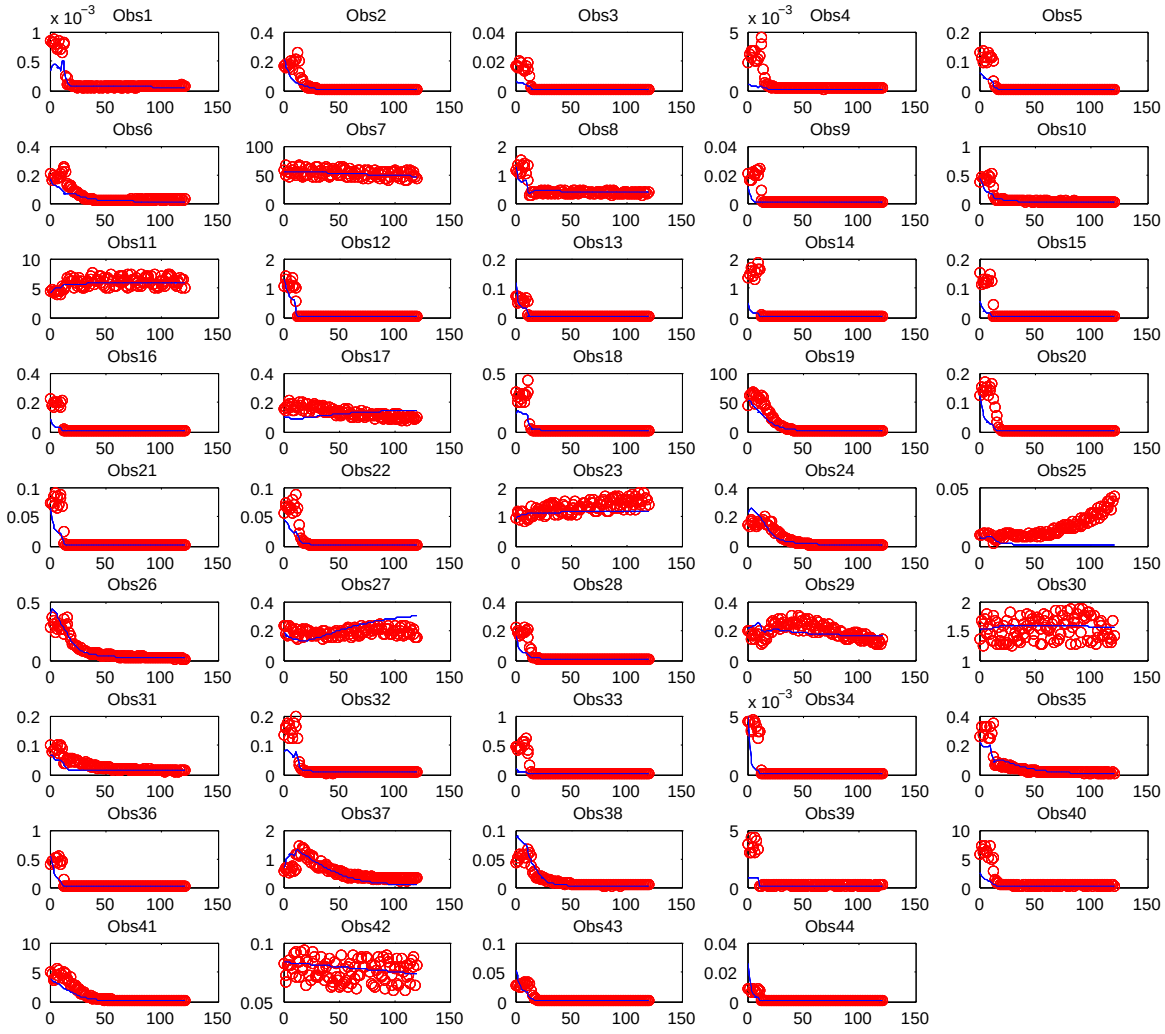

Figure A.1: **(Data fits for Yeast Model (B1))** Pseudo-experimental data (red circles) vs. optimal solution obtained using DOPS (solid blue lines) for the 44 observed states. X axis: time [s]; Y axis: metabolite concentrations [mM].

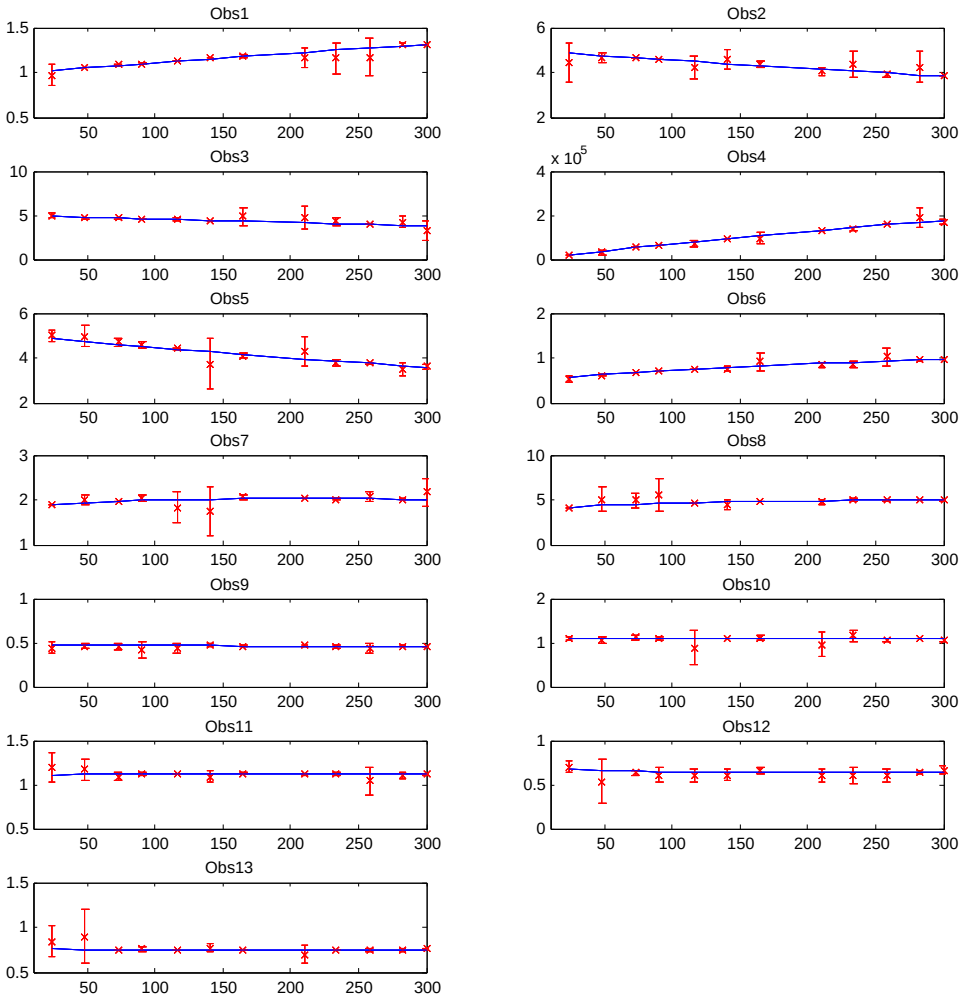

Figure A.2: **(Data fits for CHO Model (B4))** Pseudo-experimental data (red x) vs. optimal solution obtained using DOPS (solid blue lines) for the 13 observed states. X axis: time [s]; Y axis: metabolite concentrations [mM].

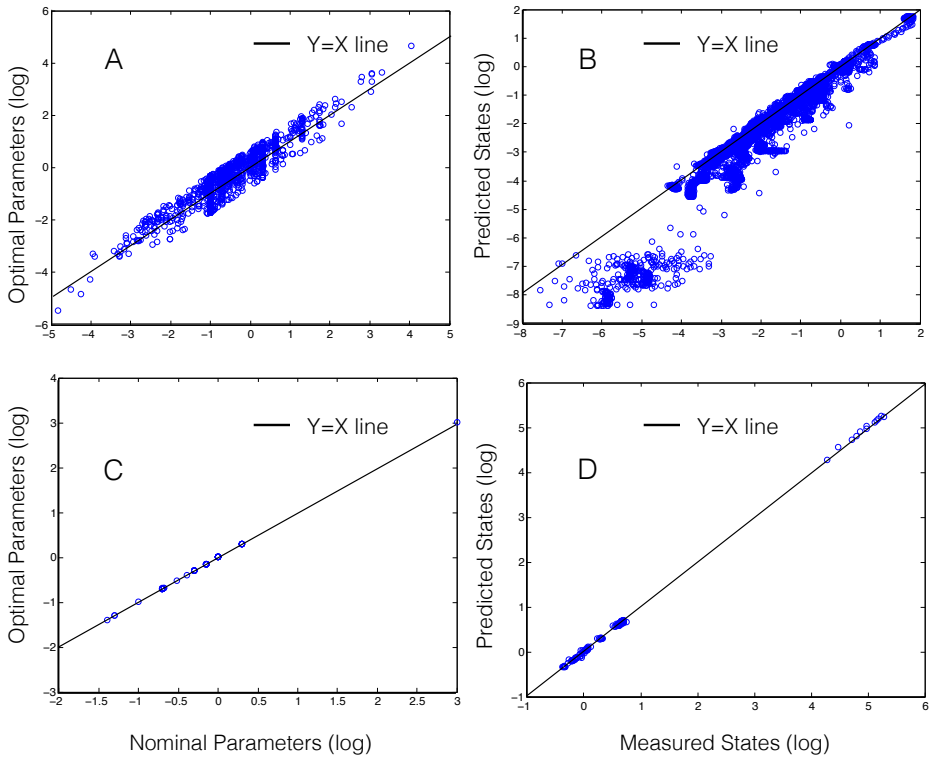

Figure A.3: **(A)** Difference between nominal and optimal parameters for problem B1: Genome wide kinetic model of *S.cerevisiae* with 1759 unknown parameters. **(B)** Difference between experimental (measured) data and data simulated with optimal parameters for yeast model: Genome wide kinetic model of *S.cerevisiae* with 1759 unknown parameters. **(C)** Difference between nominal and optimal parameters for problem B4: Metabolic model of Chinese Hamster Ovary Cells (CHO) cells with 117 parameters. **(D)** Difference between experimental (measured) data and data simulated with optimal parameters for CHO model: Metabolic model of Chinese Hamster Ovary Cells (CHO) cells with 117 parameters.

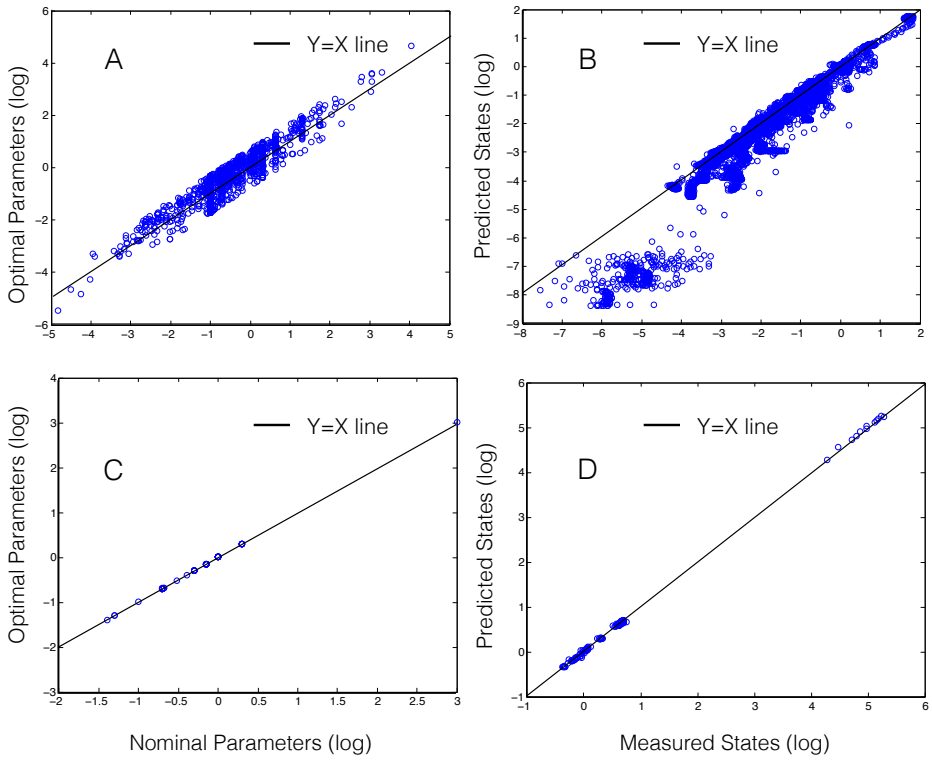

Figure A.4: **(A)** Difference between nominal and optimal parameters for problem B1: Genome wide kinetic model of *S.cerevisiae* with 1759 unknown parameters. **(B)** Difference between experimental (measured) data and data simulated with optimal parameters for yeast model: Genome wide kinetic model of *S.cerevisiae* with 1759 unknown parameters. **(C)** Difference between nominal and optimal parameters for problem B4: Metabolic model of Chinese Hamster Ovary Cells (CHO) cells with 117 parameters. **(D)** Difference between experimental (measured) data and data simulated with optimal parameters for CHO model: Metabolic model of Chinese Hamster Ovary Cells (CHO) cells with 117 parameters.

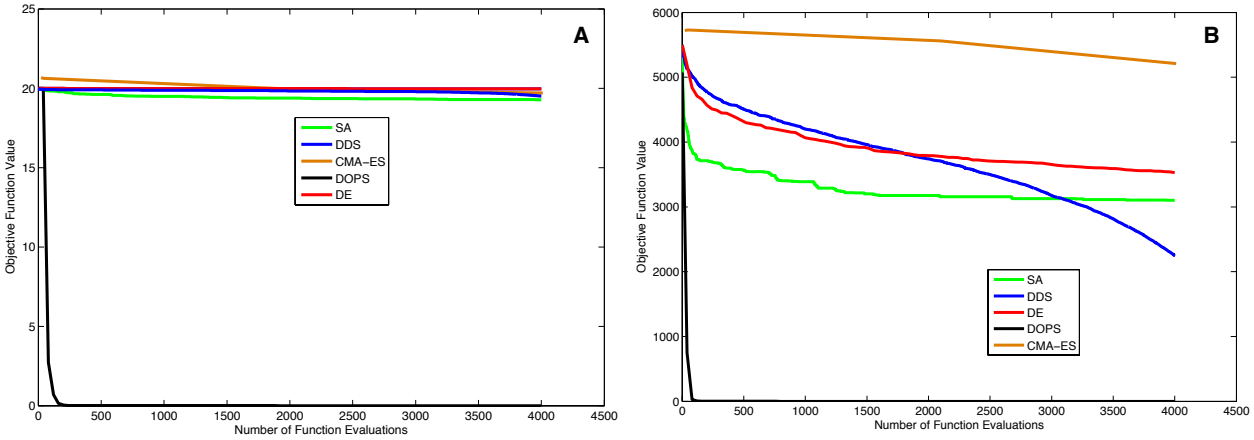

Figure A.5: Convergence curves for different metaheuristics along with CMA-ES) Objective function value versus the number of function evaluations for **(A)** Ackley 300 dimensional **(B)** and Rastrigin 300 dimensional.

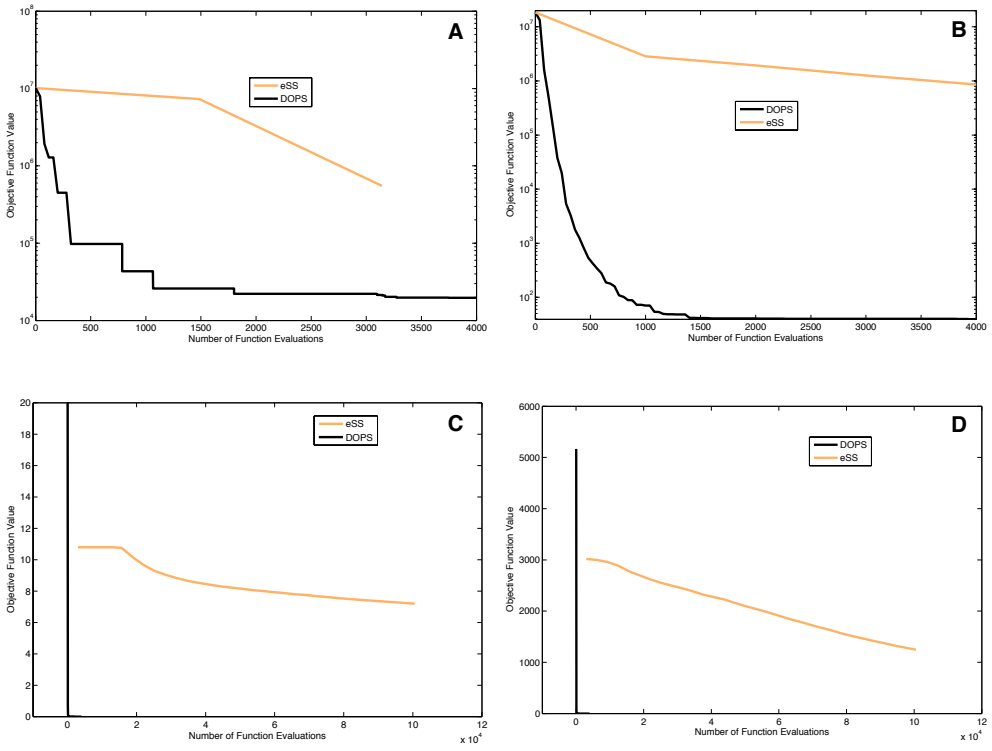

Figure A.6: (Convergence curves for eSS and DOPS) Objective function value versus the number of function evaluations plotted for eSS and DOPS on **(A)** Coagulation **(B)** CHO Model **(C)** Ackley 300 dimensional and **(D)** Rastrigin 300 dimensional.

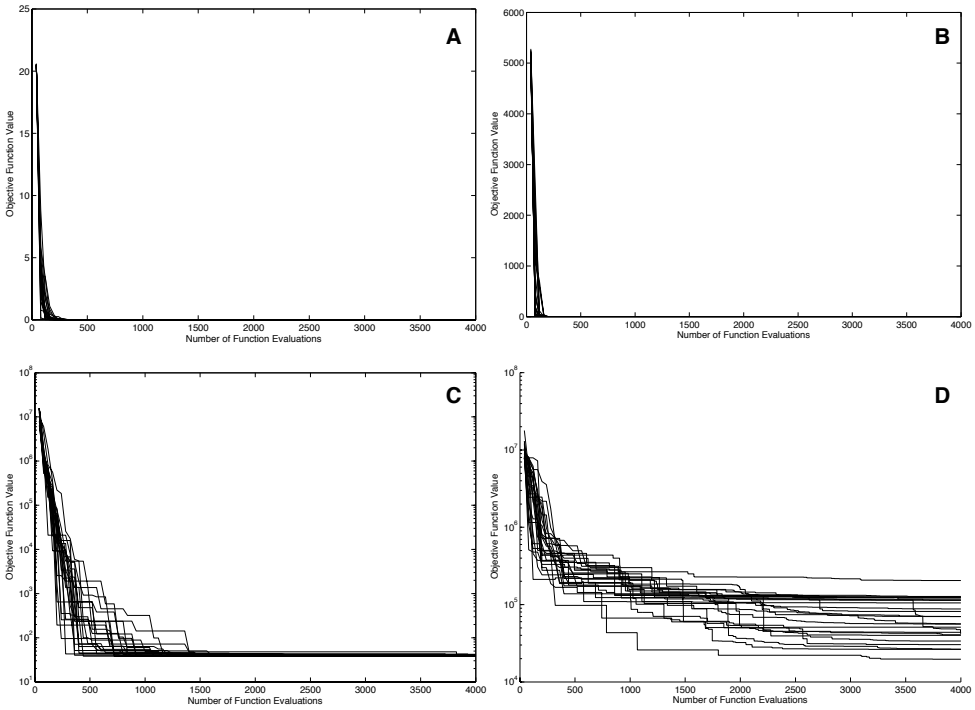

Figure A.7: (Dispersion curves for DOPS) Objective function value (25 trials) versus the number of function evaluations plotted for DOPS on **(A)** Ackley 300 dimensional **(B)** Rastrigin 300 dimensional **(C)**CHO Model - B4 **(D)** Coagulation. For CHO Model and Coagulation Y-axis is in log scale.

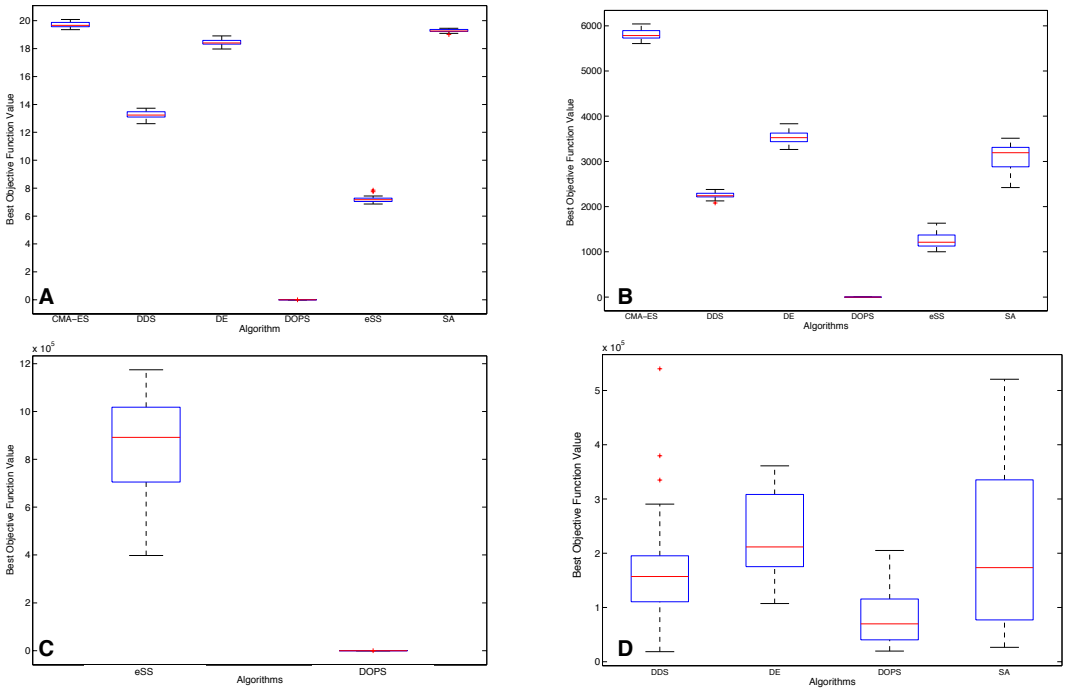

Figure A.8: Fig. S7: (Variability analysis for best objective value): Boxplots for different metaheuristics on **(A)** Ackley 300 dimensional **(B)** Rastrigin 300 dimensional **(C)** CHO Model - B4 **(D)** and Coagulation

BIBLIOGRAPHY

[1] Butenas S, van't Veer C, Mann KG (1999) "normal" thrombin generation. Blood 94: 2169-78.

[2] van't Veer C, Golden NJ, Kalafatis M, Mann KG (1997) Inhibitory mechanism of the protein c pathway on tissue factor-induced thrombin generation synergistic effect in combination with tissue factor pathway inhibitor. Journal of Biological Chemistry 272: 7983–7994.

[3] Allen G, MHoffman, Roberts H, Monroe D (2006) Manipulation of prothrombin concentration improves response to high-dose factor VIIa in a cell-based model of haemophilia. Br J Haematology 134: 314 - 319.

[4] Morad HO, Belete SC, Read T, Shaw AM (2015) Time-course analysis of c3a and c5a quantifies the coupling between the upper and terminal complement pathways in vitro. J Immunol Methods 427: 13–18.

[5] Morgan BP, Harris CL (2015) Complement, a target for therapy in inflammatory and degenerative diseases. Nat Rev Drug Discov 14: 857–877.

[6] Kholodenko B, Yaffe MB, Kolch W (2012) Computational approaches for analyzing information flow in biological networks. Sci Signal 5: re1.

[7] Machado D, Costa RS, Rocha M, Ferreira EC, Tidor B, et al. (2011) Modeling formalisms in systems biology. AMB express 1: 45.

[8] Davie EW, Fujikawa K (1975) Basic mechanisms in blood coagulation. Annual review of biochemistry 44: 799–829.

[9] Macfarlane R (1964) An enzyme cascade in the blood clotting mechanism, and its function as a biochemical amplifier. Nature 202: 498–499.

[10] Rittirsch D, Flierl MA, Ward PA (2008) Harmful molecular mechanisms in sepsis. Nat Rev Immunol 8: 776–787.

[11] Hess JR, Brohi K, Dutton RP, Hauser CJ, Holcomb JB, et al. (2008) The coagulopathy of trauma: a review of mechanisms. Journal of Trauma and Acute Care Surgery 65: 748–754.

[12] Amara U, Rittirsch D, Flierl M, Bruckner U, Klos A, et al. (2008) Interaction between the coagulation and complement system. Springer.

[13] Borkowska S, Suszynska M, Mierzejewska K, Ismail A, Budkowska M, et al. (2014) Novel evidence that crosstalk between the complement, coagulation and fibrinolysis proteolytic cascades is involved in mobilization of hematopoietic stem/progenitor cells (hspcs). Leukemia 28: 2148–2154.

[14] Markiewski MM, Nilsson B, Ekdahl KN, Mollnes TE, Lambris JD (2007) Complement and coagulation: strangers or partners in crime? Trends in immunology 28.

[15] Amara U, Flierl MA, Rittirsch D, Klos A, Chen H, et al. (2010) Molecular intercommunication between the complement and coagulation systems. The Journal of Immunology 185: 5628–5636.

[16] Janeway CA, Travers P, Walport M, Shlomchik MJ (1997) Immunobiology: the immune system in health and disease. Current Biology Singapore.

[17] Wiman B, Collen D (1978) Molecular mechanism of physiological fibrinolysis. Nature 272: 549–550.

[18] Chapin JC, Hajjar KA (2015) Fibrinolysis and the control of blood coagulation. Blood reviews 29: 17–24.

[19] Jackson G, Ashpole K, Yentis S (2009) The teg® vs the rotem® thromboelastography/thromboelastometry systems. Anaesthesia 64: 212–215.

[20] Lang T, Bauters A, Braun SL, Pötzsch B, von Pape KW, et al. (2005) Multi-centre investigation on reference ranges for rotem thromboelastometry. Blood coagulation & fibrinolysis 16: 301–310.

[21] Tolson BA, Shoemaker CA (2007) Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Resources Research 43.

[22] Kennedy J, Eberhart R (1995) Particle swarm optimization. In: Proceedings of the International Conference on Neural Networks. pp. 1942 - 1948.

[23] Gadkar KG, Varner J, Doyle FJ (2005) Model identification of signal transduction networks from data using a state regulator problem. Syst Biol (Stevenage) 2: 17–30.

[24] Bailey JE (2001) Complex biology with no parameters. Nat Biotechnol 19: 503-4.

[25] Machta BB, Chachra R, Transtrum MK, Sethna JP (2013) Parameter space compression underlies emergent theories and predictive models. Science 342: 604-7.

[26] Tasseff R, Nayak S, Song SO, Yen A, Varner JD (2011) Modeling and analysis of retinoic acid induced differentiation of uncommitted precursor cells. Integr Biol (Camb) 3: 578-91.

[27] Morris MK, Saez-Rodriguez J, Sorger PK, Lauffenburger DA (2010) Logic-based models for the analysis of cell signaling networks. Biochemistry 49: 3216-24.

[28] Saez-Rodriguez J, Alexopoulos LG, Zhang M, Morris MK, Lauffenburger DA, et al. (2011) Comparing signaling networks between normal and transformed hepatocytes using discrete logical models. Cancer Res 71: 5400-11.

[29] Morris MK, Saez-Rodriguez J, Clarke DC, Sorger PK, Lauffenburger DA (2011) Training signaling pathway maps to biochemical data with constrained fuzzy logic: quantitative analysis of liver cell responses to inflammatory stimuli. PLoS Comput Biol 7: e1001099.

[30] Morris MK, Shriver Z, Sasisekharan R, Lauffenburger DA (2012) Querying quantitative logic models (q2lm) to study intracellular signaling networks and cell-cytokine interactions. Biotechnol J 7: 374-86.

[31] Butenas S, Mann KG (2002) Blood coagulation. Biochemistry (Mosc) 67: 3-12.

[32] Schenone M, Furie BC, Furie B (2004) The blood coagulation cascade. Curr Opin Hematol 11: 272-7.

[33] Adams RLC, Bird RJ (2009) Review article: Coagulation cascade and therapeutics update: relevance to nephrology. part 1: Overview of coagulation, thrombophilias and history of anticoagulants. Nephrology (Carlton) 14: 462-70.

[34] Goldhaber SZ, Colman RW, Clowes AW, editors (2006) Hemostasis and Thrombosis: Basic Principles and Clinical Practice. Lippincott Williams and Wilkins.

[35] Brummel KE, Paradis SG, Butenas S, Mann KG (2002) Thrombin functions during tissue factor-induced blood coagulation. Blood 100: 148-52.

[36] Mann K, Nesheim M, Church W, Haley P, Krishnaswamy S (1990) Surface-dependent reactions of vitamin k-dependent enzyme complexes. Blood 76: 1-16.

[37] Roberts H, Monroe D, Oliver J, Chang J, Hoffman M (1998) Newer concepts of blood coagulation. Haemophilia 4: 331-334.

[38] Mann K (1999) Biochemistry and physiology of blood coagulation. Thromb Haemost 82: 165-174.

[39] Khanin MA, Semenov VV (1989) A mathematical model of the kinetics of blood coagulation. J Theor Biol 136: 127-34.

[40] Willems GM, Lindhout T, Hermens WT, Hemker HC (1991) Simulation model for thrombin generation in plasma. Haemostasis 21: 197-207.

[41] Baldwin SA, Basmadjian D (1994) A mathematical model of thrombin production in blood coagulation, part i: The sparsely covered membrane case. Ann Biomed Eng 22: 357-70.

[42] Leipold RJ, Bozarth TA, Racanelli AL, Dicker IB (1995) Mathematical model of serine protease inhibition in the tissue factor pathway to thrombin. J Biol Chem 270: 25383-7.

[43] Kuharsky AL, Fogelson AL (2001) Surface-mediated control of blood coagulation: the role of binding site densities and platelet deposition. Biophys J 80: 1050-74.

[44] Luan D, Zai M, Varner JD (2007) Computationally derived points of fragility of a human cascade are consistent with current therapeutic strategies. PLoS Comput Biol 3: e142.

[45] Luan D, Szlam F, Tanaka KA, Barie PS, Varner JD (2010) Ensembles of uncertain mathematical models can identify network response to therapeutic interventions. Mol Biosyst 6: 2272-86.

[46] Jones KC, Mann KG (1994) A model for the tissue factor pathway to thrombin. ii. a mathematical simulation. J Biol Chem 269: 23367-73.

[47] Hockin MF, Jones KC, Everse SJ, Mann KG (2002) A model for the stoichiometric regulation of blood coagulation. J Biol Chem 277: 18322-33.

[48] Butenas S, Orfeo T, Gissel MT, Brummel KE, Mann KG (2004) The significance of circulating factor ixa in blood. J Biol Chem 279: 22875-82.

[49] Papadopoulos KP, Gavaises M, Atkin C (2014) A simplified mathematical model for thrombin generation. Medical engineering & physics 36: 196–204.

[50] Lo K, Denney WS, Diamond SL (2005) Stochastic modeling of blood coagulation initiation. Pathophysiol Haemost Thromb 34: 80-90.

[51] Chatterjee MS, Denney WS, Jing H, Diamond SL (2010) Systems biology of coagulation initiation: kinetics of thrombin generation in resting and activated human blood. PLoS Comput Biol 6.

[52] Stalker TJ, Traxler EA, Wu J, Wannemacher KM, Cermignano SL, et al. (2013) Hierarchical organization in the hemostatic response and its relationship to the platelet-signaling network. Blood 121: 1875-85.

[53] Leiderman K, Fogelson A (2014) An overview of mathematical modeling of thrombus formation under flow. Thromb Res 133 Suppl 1: S12-4.

[54] Bannish BE, Keener JP, Fogelson AL (2014) Modelling fibrinolysis: a 3d stochastic multiscale model. Math Med Biol 31: 17-44.

[55] Voronov RS, Stalker TJ, Brass LF, Diamond SL (2013) Simulation of intrathrombus fluid and solute transport using in vivo clot structures with single platelet resolution. Ann Biomed Eng 41: 1297-307.

[56] Runyon MK, Johnson-Kerner BL, Ismagilov RF (2004) Minimal functional model of hemostasis in a biomimetic microfluidic system. Angew Chem Int Ed Engl 43: 1531-6.

[57] Kastrup CJ, Runyon MK, Shen F, Ismagilov RF (2006) Modular chemical mechanism predicts spatiotemporal dynamics of initiation in the complex network of hemostasis. Proc Natl Acad Sci U S A 103: 15747-52.

[58] Runyon MK, Johnson-Kerner BL, Kastrup CJ, Van Ha TG, Ismagilov RF (2007) Propagation of blood clotting in the complex biochemical network of hemostasis is described by a simple mechanism. J Am Chem Soc 129: 7014-5.

[59] Runyon MK, Kastrup CJ, Johnson-Kerner BL, Ha TGV, Ismagilov RF (2008) Effects of shear rate on propagation of blood clotting determined using microfluidics and numerical simulations. J Am Chem Soc 130: 3458-64.

[60] Kuepfer L, Peter M, Sauer U, Stelling J (2007) Ensemble modeling for analysis of cell signaling dynamics. Nat Biotechnol 25: 1001-6.

[61] Song SO, Varner J (2009) Modeling and analysis of the molecular basis of pain in sensory neurons. PLoS One 4: e6758.

[62] Song SO, Chakrabarti A, Varner JD (2010) Ensembles of signal transduction models using pareto optimal ensemble techniques (poets). Biotechnol J 5: 768-80.

[63] Tasseff R, Nayak S, Salim S, Kaushik P, Rizvi N, et al. (2010) Analysis of the molecular networks in androgen dependent and independent prostate cancer revealed fragile and robust subsystems. PLoS One 5: e8864.

[64] Lequieu J, Chakrabarti A, Nayak S, Varner JD (2011) Computational modeling and

analysis of insulin induced eukaryotic translation initiation. PLoS Comput Biol 7: e1002263.

[65] Tran LM, Rizk ML, Liao JC (2008) Ensemble modeling of metabolic networks. Biophys J 95: 5606-17.

[66] GKHansson (2005) Inflammation, Atherosclerosis and Coronary Artery Disease. N Engl J Med 352: 1685 - 1695.

[67] Tanaka KA, Key NS, Levy JH (2009) Blood coagulation: hemostasis and thrombin regulation. Anesth Analg 108: 1433-46.

[68] Tuddenham E, Cooper D (1994) The molecular genetics of haemostasis and its inherited disorders., volume 25 of *Oxford monographs in medical genetics*. Oxford University Press.

[69] Mannucci MP, Tuddenham EGD (2001) The hemophilias - from royal genes to gene therapy. N Engl J Med 344: 1773 - 1780.

[70] Mitchell J, Phillott A (2008) Haemophilia and inhibitors 1: diagnosis and treatment. Nursing Times 104: 26–27.

[71] Tomokiyo K, Nakatomi Y, Araki T, Teshima K, Nakano H, et al. (2003) A novel therapeutic approach combining human plasma-derived factors viia and x for haemophiliacs with inhibitors: evidence of a higher thrombin generation rate in vitro and more sustained haemostatic activity in vivo than obtained with factor viia alone. Vox Sanguinis 85: 290-299.

[72] Hedner U (2008) Factor viia and its potential therapeutic use in bleeding-associated pathologies. Thromb Haemost 100: 557–562.

[73] Talbot M, Tien HC (2009) The use of recombinant factor viia in trauma patients. J Am Acad Orthop Surg 17: 477-81.

[74] Shapiro AD (2008) Single-dose recombinant activated factor vii for the treatment of joint bleeds in hemophilia patients with inhibitors. Clin Adv Hematol Oncol 6: 579–586.

[75] Duchesne JC, Mathew KA, Marr AB, Pinsky MR, Barbeau JM, et al. (2008) Current evidence based guidelines for factor viia use in trauma: the good, the bad, and the ugly. Am Surg 74: 1159-65.

[76] Butenas S, Brummel KE, Branda RF, Paradis SG, Mann KG (2002) Mechanism of factor viia-dependent coagulation in hemophilia blood. Blood 99: 923-30.

[77] Danforth CM, Orfeo T, Everse SJ, Mann KG, Brummel-Ziedins KE (2012) Defining the boundaries of normal thrombin generation: investigations into hemostasis. PLoS One 7: e30385.

[78] Brummel-Ziedins KE, Orfeo T, Callas PW, Gissel M, Mann KG, et al. (2012) The prothrombotic phenotypes in familial protein c deficiency are differentiated by computational modeling of thrombin generation. PloS one 7: e44378.

[79] Mitrophanov AY, Wolberg AS, Reifman J (2014) Kinetic model facilitates analysis of fibrin generation and its modulation by clotting factors: implications for hemostasis-enhancing therapies. Molecular BioSystems 10: 2347–2357.

[80] Naito K, Fujikawa K (1991) Activation of human blood coagulation factor XI independent of factor XII. J Biol Chem 266: 7353-7358.

[81] Gailani D, Broze GJ Jr (1991) Factor xi activation in a revised model of blood coagulation. Science 253: 909-12.

[82] Smith SA, Mutch NJ, Baskar D, Rohloff P, Docampo R, et al. (2006) Polyphosphate modulates blood coagulation and fibrinolysis. Proc Natl Acad Sci U S A 103: 903-8.

[83] Jones E, Oliphant T, Peterson P (2001–). SciPy: Open source scientific tools for Python. http://www.scipy.org/.

[84] Hunter JD (2007) Matplotlib: A 2d graphics environment. Computing in Science and Engineering 9: 90 - 95.

[85] Sobol I (2001) Global sensitivity indices for nonlinear mathematical models and their Monte Carlo estimates. Math Comput Simulat 55: 271 - 280.

[86] Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, et al. (2010) Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. Computer Physics Communications 181: 259 - 270.

[87] Herman JD. http://jdherman.github.io/salib/.

[88] Fredrickson AG (1976) Formulation of structured growth models. Biotechnol Bioeng 18: 1481-6.

[89] Domach MM, Leung SK, Cahn RE, Cocks GG, Shuler ML (1984) Computer model for glucose-limited growth of a single cell of escherichia coli b/r-a. Biotechnol Bioeng 26: 203-16.

[90] Steinmeyer D, Shuler M (1989) Structured model for Saccharomyces cerevisiae. Chem Eng Sci 44: 2017 - 2030.

[91] Wu P, Ray NG, Shuler ML (1992) A single-cell model for cho cells. Ann N Y Acad Sci 665: 152-87.

[92] Castellanos M, Wilson DB, Shuler ML (2004) A modular minimal cell model: purine and pyrimidine transport and metabolism. Proc Natl Acad Sci U S A 101: 6681-6.

[93] Atlas JC, Nikolaev EV, Browning ST, Shuler ML (2008) Incorporating genome-wide dna sequence information into a dynamic whole-cell model of escherichia coli: application to dna replication. IET Syst Biol 2: 369-82.

[94] Dhurjati P, Ramkrishna D, Flickinger MC, Tsao GT (1985) A cybernetic view of microbial growth: modeling of cells as optimal strategists. Biotechnol Bioeng 27: 1-9.

[95] Kompala DS, Ramkrishna D, Jansen NB, Tsao GT (1986) Investigation of bacterial growth on mixed substrates: experimental evaluation of cybernetic models. Biotechnol Bioeng 28: 1044-55.

[96] Kim JI, Song HS, Sunkara SR, Lali A, Ramkrishna D (2012) Exacting predictions by cybernetic model confirmed experimentally: steady state multiplicity in the chemostat. Biotechnol Prog 28: 1160-6.

[97] Varner J, Ramkrishna D (1999) Metabolic engineering from a cybernetic perspective: aspartate family of amino acids. Metab Eng 1: 88-116.

[98] Song HS, Morgan JA, Ramkrishna D (2009) Systematic development of hybrid cybernetic models: application to recombinant yeast co-consuming glucose and xylose. Biotechnol Bioeng 103: 984-1002.

[99] Song HS, Ramkrishna D (2011) Cybernetic models based on lumped elementary modes accurately predict strain-specific metabolic function. Biotechnol Bioeng 108: 127-40.

[100] Lewis NE, Nagarajan H, Palsson BO (2012) Constraining the metabolic genotype-

phenotype relationship using a phylogeny of in silico methods. Nat Rev Microbiol 10: 291-305.

[101] Edwards JS, Palsson BO (2000) The escherichia coli mg1655 in silico metabolic genotype: its definition, characteristics, and capabilities. Proc Natl Acad Sci U S A 97: 5528-33.

[102] Feist AM, Henry CS, Reed JL, Krummenacker M, Joyce AR, et al. (2007) A genome-scale metabolic reconstruction for escherichia coli k-12 mg1655 that accounts for 1260 orfs and thermodynamic information. Mol Syst Biol 3: 121.

[103] Oh YK, Palsson BO, Park SM, Schilling CH, Mahadevan R (2007) Genome-scale reconstruction of metabolic network in bacillus subtilis based on high-throughput phenotyping and gene essentiality data. J Biol Chem 282: 28791-9.

[104] Feist AM, Herrgrd MJ, Thiele I, Reed JL, Palsson BØ (2009) Reconstruction of biochemical networks in microorganisms. Nat Rev Microbiol 7: 129-43.

[105] Ibarra RU, Edwards JS, Palsson BO (2002) Escherichia coli k-12 undergoes adaptive evolution to achieve in silico predicted optimal growth. Nature 420: 186-9.

[106] Schuetz R, Kuepfer L, Sauer U (2007) Systematic evaluation of objective functions for predicting intracellular fluxes in escherichia coli. Mol Syst Biol 3: 119.

[107] Hyduke DR, Lewis NE, Palsson BØ (2013) Analysis of omics data with genome-scale models of metabolism. Mol Biosyst 9: 167-74.

[108] McCloskey D, Palsson BØ, Feist AM (2013) Basic and applied uses of genome-scale metabolic network reconstructions of escherichia coli. Mol Syst Biol 9: 661.

[109] Zomorrodi AR, Suthers PF, Ranganathan S, Maranas CD (2012) Mathematical optimization applications in metabolic networks. Metab Eng 14: 672-86.

[110] Jewett MC, Calhoun KA, Voloshin A, Wuu JJ, Swartz JR (2008) An integrated cell-free metabolic platform for protein production and synthetic biology. Mol Syst Biol 4: 220.

[111] MATTHAEI JH, NIRENBERG MW (1961) Characteristics and stabilization of dnaase-sensitive protein synthesis in e. coli extracts. Proc Natl Acad Sci U S A 47: 1580-8.

[112] NIRENBERG MW, MATTHAEI JH (1961) The dependence of cell-free protein synthesis in e. coli upon naturally occurring or synthetic polyribonucleotides. Proc Natl Acad Sci U S A 47: 1588-602.

[113] Lu Y, Welsh JP, Swartz JR (2014) Production and stabilization of the trimeric influenza hemagglutinin stem domain for potentially broadly protective influenza vaccines. Proc Natl Acad Sci U S A 111: 125-30.

[114] Hodgman CE, Jewett MC (2012) Cell-free synthetic biology: thinking outside the cell. Metab Eng 14: 261-9.

[115] Morris MK, Saez-Rodriguez J, Clarke DC, Sorger PK, Lauffenburger DA (2011) Training signaling pathway maps to biochemical data with constrained fuzzy logic: quantitative analysis of liver cell responses to inflammatory stimuli. PLoS Comput Biol 7: e1001099.

[116] Brown KS, Sethna JP (2003) Statistical mechanical approaches to models with many poorly known parameters. Phys Rev E Stat Nonlin Soft Matter Phys 68: 021904.

[117] Covert MW, Schilling CH, Palsson B (2001) Regulation of gene expression in flux balance models of metabolism. J Theor Biol 213: 73-88.

[118] Covert MW, Knight EM, Reed JL, Herrgard MJ, Palsson BO (2004) Integrating high-throughput and computational data elucidates bacterial networks. Nature 429: 92-6.

[119] Varner JD (2000) Large-scale prediction of phenotype: concept. Biotechnol Bioeng 69: 664-78.

[120] Song HS, Ramkrishna D (2012) Prediction of dynamic behavior of mutant strains from limited wild-type data. Metab Eng 14: 69-80.

[121] Gadkar KG, Doyle FJ 3rd, Crowley TJ, Varner JD (2003) Cybernetic model predictive control of a continuous bioreactor with cell recycle. Biotechnol Prog 19: 1487-97.

[122] Nayak S, Siddiqui JK, Varner JD (2011) Modelling and analysis of an ensemble of eukaryotic translation initiation models. IET Syst Biol 5: 2.

[123] Berg JM, Tymoczko JL, Stryer L (2002) Biochemistry. W.H. Freeman.

[124] Peskov K, Goryanin I, Demin O (2008) Kinetic model of phosphofructokinase-1 from escherichia coli. J Bioinform Comput Biol 6: 843-67.

[125] Keseler IM, Mackie A, Peralta-Gil M, Santos-Zavaleta A, Gama-Castro S, et al. (2013) Ecocyc: fusing model organism databases with systems biology. Nucleic Acids Res 41: D605-12.

[126] Huang Z, Mou L, Shen Q, Lu S, Li C, et al. (2014) Asd v2.0: updated content and novel features focusing on allosteric regulation. Nucleic Acids Res 42: D510-6.

[127] Link H, Kochanowski K, Sauer U (2013) Systematic identification of allosteric protein-metabolite interactions that control enzyme activity in vivo. Nat Biotechnol 31: 357-61.

[128] Kremling A, Fischer S, Gadkar K, Doyle FJ, Sauter T, et al. (2004) A benchmark for methods in reverse engineering and model discrimination: problem formulation and solutions. Genome Res 14: 1773-85.

[129] Gadkar KG, Gunawan R, Doyle FJ 3rd (2005) Iterative approach to model identification of biological networks. BMC Bioinformatics 6: 155.

[130] Song SO, Chakrabarti A, Varner JD (2010) Ensembles of signal transduction models using pareto optimal ensemble techniques (poets). Biotechnol J 5: 768-80.

[131] Nuttall G (1888) Experimente über die bacterienfeindlichen Einflüsse des thierischen Körpers. Z Hyg Infektionskr 4: 353-394.

[132] Ricklin D, Hajishengallis G, Yang K, Lambris JD (2010) Complement: a key system for immune surveillance and homeostasis. Nat Immunol 11: 785–797.

[133] Ricklin D, Lambris JD (2007) Complement-targeted therapeutics. Nat Biotechnol 25: 1265-1275.

[134] Sarma JV, Ward PA (2011) The complement system. Cell Tissue Res 343: 227–235.

[135] Ricklin D, Lambris JD (2013) Complement in immune and inflammatory disorders: pathophysiological mechanisms. J Immunol 190: 3831–3838.

[136] Ricklin D, Lambris JD (2013) Progress and trends in complement therapeutics. Adv Exp Med Biol 735: 1-22.

[137] Walport MJ (2001) Complement. first of two parts. N Engl J Med 344: 1058-66.

[138] Walport MJ (2001) Complement. second of two parts. N Engl J Med 344: 1140-4.

[139] Pangburn MK, Müller-Eberhard HJ (1984) The alternative pathway of complement. Springer Semin Immunopathol 7: 163–192.

[140] Walker D, Yasuhara O, Patston P, McGeer E, McGeer P (1995) Complement c1 inhibitor is produced by brain tissue and is cleaved in alzheimer disease. Brain Res 675: 75–82.

[141] Blom AM, Kask L, Dahlbäck B (2001) Structural requirements for the complement regulatory activities of c4bp. J Biol Chem 276: 27136–27144.

[142] Riley-Vargas RC, Gill DB, Kemper C, Liszewski MK, Atkinson JP (2004) Cd46: expanding beyond complement regulation. Trends Immunol 25: 496–503.

[143] Lukacik P, Roversi P, White J, Esser D, Smith G, et al. (2004) Complement regulation at the molecular level: the structure of decay-accelerating factor. Proc Natl Acad Sci USA 101: 1279–1284.

[144] Liszewski MK, Farries TC, Lublin DM, Rooney IA, Atkinson JP (1995) Control of the complement system. Adv Immunol 61: 201–283.

[145] Chauhan A, Moore T (2006) Presence of plasma complement regulatory proteins clusterin (apo j) and vitronectin (s40) on circulating immune complexes (cic). Clin Exp Immunol 145: 398–406.

[146] Zewde N, Gorham Jr RD, Dorado A, Morikis D (2016) Quantitative modeling of the alternative pathway of the complement system. PloS One 11: e0152337.

[147] Hirayama H, Yoshii K, Ojima H, Kawai N, Gotoh S, et al. (1996) Linear systems analysis of activating processes of complement system as a defense mechanism. Biosystems 39: 173–185.

[148] Korotaevskiy AA, Hanin LG, Khanin MA (2009) Non-linear dynamics of the complement system activation. Math Biosci 222: 127–143.

[149] Liu B, Zhang J, Tan PY, Hsu D, Blom AM, et al. (2011) A computational and experimental study of the regulatory mechanisms of the complement system. PLoS Comput Biol 7: e1001059.

[150] Bassen D, Vilkhovoy M, Minot M, Butcher JT, Varner JD (2016) JuPOETs: A Constrained Multiobjective Optimization Approach to Estimate Biochemical Model Ensembles in the Julia Programming Language. bioRxiv 10.1101/056044.

[151] Nayak S, Salim S, Luan D, Zai M, Varner JD (2008) A test of highly optimized tolerance reveals fragile cell-cycle mechanisms are molecular targets in clinical cancer trials. PLoS One 3: e2016.

[152] Rice NT, Szlam F, Varner JD, Bernstein PS, Szlam AD, et al. (2016) Differential contributions of intrinsic and extrinsic pathways to thrombin generation in adult, maternal and cord plasma samples. PLoS One 11: e0154127.

[153] Schwaeble HW, Stover CM, Tedford CE, Parent JB, Fujita T (2011). Methods for treating conditions associated with masp-2 dependent complement activation. US Patent 7,919,094.

[154] Vogel CW, Fritzinger DC, Hew BE, Thorne M, Bammert H (2004) Recombinant cobra venom factor. Molecular immunology 41: 191–199.

[155] Katschke KJ, Wu P, Ganesan R, Kelley RF, Mathieu MA, et al. (2012) Inhibiting alternative pathway complement activation by targeting the factor d exosite. J Biol Chem 287: 12886–12892.

[156] Hu X, Holers VM, Thurman JM, Schoeb TR, Ramos TN, et al. (2013) Therapeutic inhibition of the alternative complement pathway attenuates chronic EAE. Mol Immunol 54: 302–308.

[157] Bansal R (2014). Humanized and chimeric anti-properdin antibodies. US Patent 8,664,362.

[158] Roguska M, Splawski I, Diefenbach-Streiber B, Dolan E, Etemad-Gilbertson B, et al. (2014) Generation and Characterization of LFG316, A Fully-Human Anti-C5 Antibody for the Treatment of Age-Related Macular Degeneration. IOVS 55: 3433–3433.

[159] Melis JP, Strumane K, Ruuls SR, Beurskens FJ, Schuurman J, et al. (2015) Complement in therapy and disease: Regulating the complement system with antibody-based therapeutics. Mol Immunol 67: 117–130.

[160] Weston-Davies WH, Nunn MA, Pinto FO, Mackie IJ, Richards SJ, et al. (2014) Clinical and immunological characterisation of coversin, a novel small protein inhibitor of complement C5 with potential as a therapeutic agent in PNH and other complement mediated disorders. Blood 124: 4280–4280.

[161] Epstein D, Kurz JC (2007). Complement binding aptamers and anti-c5 agents useful in the treatment of ocular disorders. US Patent App. 12/224,708.

[162] Mastellos DC, Yancopoulou D, Kokkinos P, Huber-Lang M, Hajishengallis G, et al. (2015) Compstatin: a c3-targeted complement inhibitor reaching its prime for bedside intervention. Eur J Clin Invest 45: 423-40.

[163] Sissons J, Liebowitch J, Amos N, Peters D (1977) Metabolism of the fifth component of complement, and its relation to metabolism of the third component, in patients with complement activation. J Clin Invest 59: 704.

[164] Swaak A, Hannema A, Vogelaar C, Boom F, van Es L, et al. (1982) Determination of the half-life of c3 in patients and its relation to the presence of c3-breakdown products and/or circulating immune complexes. Rheumatol Int : 161–166.

[165] Noris M, Galbusera M, Gastoldi S, Macor P, Banterla F, et al. (2014) Dynamics of complement activation in ahus and how to monitor eculizumab therapy. Blood : 1715–1726.

[166] Sagar A, Varner JD (2015) Dynamic modeling of the human coagulation cascade using reduced order effective kinetic models. Processes 3: 178.

[167] Bezanson J, Edelman A, Karpinski S, Shah VB (2015) Julia: A fresh approach to numerical computing. arXiv arXiv:1411.1607v4.

[168] Hindmarsh AC, Brown PN, Grant KE, Lee SL, Serban R, et al. (2005) Sundials: Suite of nonlinear and differential/algebraic equation solvers. ACM Trans Math Softw 31: 363–396.

[169] Varnerlab. http://www.varnerlab.org.

[170] Sagar A, Shoemaker CA, Varner J (2016) Dynamic Optimization with Particle Swarms (DOPS): A meta- heuristic for parameter estimation in biochemical models. Biotechnol J submitted.

[171] Tolson BA, Shoemaker CA (2007) Dynamically dimensioned search algorithm for computationally efficient watershed model calibration. Water Res Research 43: W01413.

[172] Varnerlab. JuPOETs: A Constrained Multiobjective Optimization Approach to Estimate Biochemical Model Ensembles in the Julia Programming Language. URL https://github.com/varnerlab/POETs.jl.

[173] Akaike H (1974) A new look at the statistical model identification. IEEE Trans Auto Cont 19: 716 - 723.

[174] Krug EG, Sharma GK, Lozano R (2000) The global burden of injuries. Am J Public Health 90: 523-6.

[175] for Disease Control C, Prevention, et al. (2012). Web-based injury statistics query and reporting system (wisqars). atlanta, ga: National center for injury prevention and control. retrieved march 15, 2012. WebPage.

[176] Sauaia A, Moore FA, Moore EE, Moser KS, Brennan R, et al. (1995) Epidemiology of trauma deaths: a reassessment. J Trauma 38: 185-93.

[177] Esmon CT (2005) The interactions between inflammation and coagulation. Br J Haematol 131: 417-30.

[178] Sauaia A, Moore FA, Moore EE, Haenel JB, Read RA, et al. (1994) Early predictors of postinjury multiple organ failure. Arch Surg 129: 39-45.

[179] Chapin JC, Hajjar KA (2015) Fibrinolysis and the control of blood coagulation. Blood Rev 29: 17-24.

[180] Luan D, Szlam F, Tanaka KA, Barie PS, Varner JD (2010) Ensembles of uncertain mathematical models can identify network response to therapeutic interventions. Mol Biosyst 6: 2272-86.

[181] Longstaff C, Thelwell C (2005) Understanding the enzymology of fibrinolysis and improving thrombolytic therapy. FEBS letters 579: 3303–3309.

[182] Bannish BE, Keener JP, Fogelson AL (2012) Modelling fibrinolysis: a 3d stochastic multiscale model. Mathematical Medicine and Biology : dqs029.

[183] Ahmad SS, Rawala-Sheikh R, Walsh PN (1992) Components and assembly of the factor x activating complex. In: Seminars in thrombosis and hemostasis. Copyright© 1992 by Thieme Medical Publishers, Inc., volume 18, pp. 311–323.

[184] Middleton G, Ruzevick B (2004) Alteplase (cathflo activase). Clinical journal of oncology nursing 8: 417–418.

[185] Björk I, Lindahl U (1982) Mechanism of the anticoagulant action of heparin. Molecular and cellular biochemistry 48: 161–182.

[186] Collet J, Park D, Lesty C, Soria J, Soria C, et al. (2000) Influence of fibrin network conformation and fibrin fiber diameter on fibrinolysis speed dynamic and structural approaches by confocal microscopy. Arteriosclerosis, thrombosis, and vascular biology 20: 1354–1361.

[187] Higgins DL, Mann KG (1983) The interaction of bovine factor v and factor v-derived peptides with phospholipid vesicles. J Biol Chem 258: 6503-8.

[188] Cawthern KM, van 't Veer C, Lock JB, DiLorenzo ME, Branda RF, et al. (1998) Blood coagulation in hemophilia a and hemophilia c. Blood 91: 4581-92.

[189] Williams EB, Krishnaswamy S, Mann KG (1989) Zymogen/enzyme discrimination using peptide chloromethyl ketones. J Biol Chem 264: 7536-45.

[190] Whelihan MF, Zachary V, Orfeo T, Mann KG (2012) Prothrombin activation in blood coagulation: the erythrocyte contribution to thrombin generation. Blood 120: 3837-45.

[191] Rivard GE, Brummel-Ziedins KE, Mann KG, Fan L, Hofer A, et al. (2005) Evaluation of the profile of thrombin generation during the process of whole blood clotting as assessed by thrombelastography. J Thromb Haemost 3: 2039-43.

[192] Saltelli A, Annoni P, Azzini I, Campolongo F, Ratto M, et al. (2010) Variance based sensitivity analysis of model output. design and estimator for the total sensitivity index. Comput Phys Commun 181: 259–270.

[193] Morris MD (1991) Factorial sampling plans for preliminary computational experiments. Technometrics 33: 161–174.

[194] Sumner T, Shephard E, Bogle I (2012) A methodology for global-sensitivity analysis of time-dependent outputs in systems biology modelling. Journal of The Royal Society Interface 9: 2156–2166.

[195] Campolongo F, Cariboni J, Saltelli A (2007) An effective screening design for sensitivity analysis of large models. Environmental modelling & software 22: 1509–1518.

[196] Rousseeuw PJ (1987) Silhouettes: a graphical aid to the interpretation and validation of cluster analysis. Journal of computational and applied mathematics 20: 53–65.

[197] Assmus HE, Herwig R, Cho KH, Wolkenhauer O (2006) Dynamics of biological systems: role of systems biology in medical research. Expert Review of Molecular Diagnostics .

[198] van Riel NAW (2006) Dynamic modelling and analysis of biochemical networks: mechanism-based models and model-based experiments. Brief Bioinform 7: 364-74.

[199] Jaqaman K, Danuser G (2006) Linking data to models: data regression. Nat Rev Mol Cell Biol 7: 813-9.

[200] Kitano H (2002) Systems biology: a brief overview. Science 295: 1662–1664.

[201] Hood L, Heath JR, Phelps ME, Lin B (2004) Systems biology and new technologies enable predictive and preventative medicine. Science 306: 640–643.

[202] Aldridge BB, Burke JM, Lauffenburger DA, Sorger PK (2006) Physicochemical modelling of cell signalling pathways. Nat Cell Biol 8: 1195-203.

[203] Banga JR (2008) Optimization in computational systems biology. BMC systems biology 2: 47.

[204] Ashyraliyev M, Fomekong-Nanfack Y, Kaandorp JA, Blom JG (2009) Systems biology: parameter estimation for biochemical models. Febs Journal 276: 886–902.

[205] Moles CG, Mendes P, Banga JR (2003) Parameter estimation in biochemical pathways: a comparison of global optimization methods. Genome research 13: 2467–2474.

[206] Nieman R, Fisher D, Seborg D (1971) A review of process identification and parameter estimation techniques†. International Journal of Control 13: 209–264.

[207] Beck JV, Arnold KJ (1977) Parameter estimation in engineering and science. James Beck.

[208] Young P (1981) Parameter estimation for continuous-time models—a survey. Automatica 17: 23–39.

[209] Beck JV, Woodbury KA (1998) Inverse problems and parameter estimation: integration of measurements and analysis. Measurement Science and Technology 9: 839.

[210] Hooke R, Jeeves TA (1961) "direct search"solution of numerical and statistical problems. Journal of the ACM (JACM) 8: 212–229.

[211] Nelder JA, Mead R (1965) A simplex method for function minimization. The computer journal 7: 308–313.

[212] Moré JJ (1978) The levenberg-marquardt algorithm: implementation and theory. In: Numerical analysis, Springer. pp. 105–116.

[213] Esposito WR, Floudas CA (2000) Deterministic global optimization in nonlinear optimal control problems. Journal of Global Optimization 17: 97–126.

[214] Horst R, Tuy H (2013) Global optimization: Deterministic approaches. Springer Science & Business Media.

[215] Goldberg DE (2006) Genetic algorithms. Pearson Education India .

[216] Kirkpatrick S, Gelatt CD, Vecchi MP, et al. (1983) Optimization by simulated annealing. science 220: 671–680.

[217] Fogel D (2009) Artificial intelligence through simulated evolution. Wiley-IEEE Press .

[218] Storn R, Price K (1997) Differential evolution–a simple and efficient heuristic for global optimization over continuous spaces. Journal of global optimization 11: 341–359.

[219] Tsai KY, Wang FS (2005) Evolutionary optimization with data collocation for reverse engineering of biological networks. Bioinformatics 21: 1180–1188.

[220] Wang FS, Su TL, Jang HJ (2001) Hybrid differential evolution for problems of kinetic parameter estimation and dynamic optimization of an ethanol fermentation process. Industrial & engineering chemistry research 40: 2876–2885.

[221] Noman N, Iba H (2007) Inferring gene regulatory networks using differential evolution with local search heuristics. IEEE/ACM Transactions on Computational Biology and Bioinformatics (TCBB) 4: 634–647.

[222] Sun J, Garibaldi JM, Hodgman C (2012) Parameter estimation using metaheuristics in systems biology: a comprehensive review. Computational Biology and Bioinformatics, IEEE/ACM Transactions on 9: 185–202.

[223] Mendes P, Kell D (1998) Non-linear optimization of biochemical pathways: applications to metabolic engineering and parameter estimation. Bioinformatics 14: 869–883.

[224] Modchang C, Triampo W, Lenbury Y (2008) Mathematical modeling and application of genetic algorithm to parameter estimation in signal transduction: Trafficking and promiscuous coupling of g-protein coupled receptors. Computers in Biology and Medicine 38: 574–582.

[225] Tashkova K, Korošec P, Šilc J, Todorovski L, Džeroski S (2011) Parameter estimation with bio-inspired meta-heuristic optimization: modeling the dynamics of endocytosis. BMC systems biology 5: 159.

[226] Villaverde AF, Egea JA, Banga JR (2012) A cooperative strategy for parameter estimation in large scale systems biology models. BMC systems biology 6: 75.

[227] Rodriguez-Fernandez M, Egea JA, Banga JR (2006) Novel metaheuristic for parameter estimation in nonlinear dynamic biological systems. BMC bioinformatics 7: 483.

[228] Egea JA, Rodríguez-Fernández M, Banga JR, Martí R (2007) Scatter search for chemical and bio-process optimization. Journal of Global Optimization 37: 481–503.

[229] Egea JA, Martí R, Banga JR (2010) An evolutionary method for complex-process optimization. Computers & Operations Research 37: 315–324.

[230] Villaverde AF, Henriques D, Smallbone K, Bongard S, Schmid J, et al. (2015) Biopredyn-bench: a suite of benchmark problems for dynamic modelling in systems biology. BMC systems biology 9: 8.

[231] Zhao SZ, Liang JJ, Suganthan PN, Tasgetiren MF (2008) Dynamic multi-swarm particle swarm optimizer with local search for large scale global optimization. Evolutionary Computation, 2008 CEC 2008 : 3845–3852.

[232] Villaverde AF, Bongard S, Mauch K, Müller D, Balsa-Canto E, et al. (2014) High-confidence predictions in systems biology dynamic models. 8th International Conference on Practical Applications of Computational Biology & Bioinformatics (PACBB 2014) : 161–171.

[233] Smallbone K, Mendes P (2013) Large-scale metabolic models: From reconstruction to differential equations. Industrial Biotechnology 9: 179–184.

[234] Mann KG, Butenas S, Brummel K (2003) The dynamics of thrombin formation. Arteriosclerosis, thrombosis, and vascular biology 23: 17–25.

[235] Mann K, Brummel K, Butenas S (2003) What is all that thrombin for? Journal of Thrombosis and Haemostasis 1: 1504–1514.

[236] Mann KG (2003) Thrombin formation. CHEST Journal 124: 4S–10S.

[237] Vogler EA, Siedlecki CA (2009) Contact activation of blood-plasma coagulation. Biomaterials 30: 1857–1869.

[238] Diamond SL (2013) Systems biology of coagulation. Journal of Thrombosis and Haemostasis 11: 224–232.

[239] Fogelson AL, Tania N (2005) Coagulation under flow: the influence of flow-mediated transport on the initiation and inhibition of coagulation. Pathophysiology of haemostasis and thrombosis 34: 91–108.

[240] Anand M, Rajagopal K, Rajagopal K (2003) A model incorporating some of the mechanical and biochemical factors underlying clot formation and dissolution in flowing blood: review article. Journal of Theoretical Medicine 5: 183–218.

[241] Hockin MF, Jones KC, Everse SJ, Mann KG (2002) A model for the stoichiometric regulation of blood coagulation. Journal of Biological Chemistry 277: 18322–18333.

[242] Chatterjee MS, Denney WS, Jing H, Diamond SL (2010) Systems biology of coagulation initiation: kinetics of thrombin generation in resting and activated human blood. PLoS computational biology .

[243] Mann KG, Brummel-Ziedins K, Orfeo T, Butenas S (2006) Models of blood coagulation. Blood Cells, Molecules, and Diseases 36: 108–117.

[244] Luan D, Zai M, Varner JD (2007) Computationally derived points of fragility of a human cascade are consistent with current therapeutic strategies. PLoS computational biology 3: e142.

[245] Song SO, Chakrabarti A, Varner JD (2010) Ensembles of signal transduction models using pareto optimal ensemble techniques (POETs). Biotechnology journal 5: 768–780.

[246] Clerc M (2010) Particle swarm optimization, volume 93. John Wiley & Sons.

[247] Blondin J (2009) Particle swarm optimization: A tutorial. from site: http://cs armstrong edu/saad/csci8100/pso tutorial pdf 93.

[248] Abraham A, Guo H, Liu H (2006) Swarm intelligence: foundations, perspectives and applications. In: Swarm Intelligent Systems, Springer. pp. 3–25.

[249] Shi Y, Eberhart RC (1999) Empirical study of particle swarm optimization. In: Proceedings of the 1999 Congress on Evolutionary Computation, 1999. CEC 99. volume 3, p. 1950.